Πανεπιστήμιο Θεσσαλίας

Αναγνώριση συναισθήματος από βίντεο με τη χρήση νευρωνικών δικτύων

Διπλωματική εργασία στο τμήμα
Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
του

Κοσκινά Ιωάννη

Επιβλέποντες Καθηγητές:
Δρ. Ποταμιάνος Γεράσιμος
Δρ. Αργυρίου Αντώνιος

26 Ιουνίου 2018

Δήλωση Προστασίας Πνευματικών Δικαιωμάτων

Εγώ, ο Κοσκινάς Ιωάννης, δηλώνω πως η συγκεκριμένη πτυχιακή εργασία με θέμα την Αναγνώριση Συναισθημάτων από βίντεο και η δουλειά που παρουσιάζεται εδώ αποτελεί προσωπική προσπάθεια. Επιβεβαιώνω πως:

- Αυτή η εργασία υλοποιήθηκε αποκλειστικά στα πλαίσια της πτυχιακής εργασίας για το Πανεπιστήμιο Θεσσαλίας.
- Οποιοδήποτε μέρος της εργασίας που αποτελεί μέρος άλλης δημοσιευμένης εργασίας έχει δηλωθεί ξεκάθαρα στην βιβλιογραφία.

'Η φαντασία μας είναι το μόνο όριο στο τι ελπίζουμε να έχουμε στο μέλλον.'

Charles F. Kettering

Περίληψη

Αδιαμφισβήτητα, η ολοένα και μεγαλύτερη αλληλεπίδραση μεταξύ ανθρώπων και μηχανών στην καθημερινότητα καθιστά αναγκαία την καλύτερη κατανόηση των ανθρώπινων συναισθημάτων από την πλευρά των μηχανών. Σε αυτή τη Διπλωματική γίνεται μια προσπάθεια για την κατηγοριοποίηση των συναισθημάτων με βάση τα φυσικά χαρακτηριστικά του προσώπου σε μία από τις εφτά βασικές κατηγορίες: τον θυμό, την απέχθεια, τον φόβο, την χαρά, την αηδία, την λύπη και την έκπληξη με την χρήση Νευρωνικών Δικτύων Συνέλιξης (ΝΔΣ). Στα πλαίσια της εργασίας αρχικά εξετάζεται η ανάλυση διάφορων τεχνικών προεπεξεργασίας πάνω σε εικόνες, που θα δημιουργήσουν ευνοϊκές συνθήκες για την υποδοχή των δεδομένων από το ΝΔΣ. Στην συνέχεια, η έρευνα επικεντρώνεται στην κατασκευή από την αρχή και την εκπαίδευση ενός μοντέλου ΝΔΣ ικανού να κατηγοριοποιήσει αποδοτικά τα συναισθήματα. Τέλος, μέσα από τα διάφορα πειράματα που έχουν πραγματοποιηθεί εξάγονται συμπεράσματα και απεικονίζονται αναλυτικά τα αποτελέσματα.

Abstract

Undoubtedly, the growing interaction between people and machines in everyday life necessitates a better understanding of human feelings on the part of machines. In this Thesis, an attempt has been made to categorize emotions based on the physical characteristics of the person in one of the seven main categories: anger, contempt, fear, happiness, disgust, sadness, and surprise using Convolutional Neural Networks. This Thesis contains analysis of various preprocessing techniques applied on images in order to prepare the data as an appropriate input to the Convolutional Neural Network. The research then focuses on the creation of a Convolutional Neural Network model from scratch capable of classifying emotions efficiently. Finally, all results are analyzed and examined thoroughly.

Ευχαριστίες

Στο συγκεκριμένο σημείο αισθάνομαι την υποχρέωση να εκφράσω κάποιες ευχαριστίες σε ορισμένους ανθρώπους που συνεργάστηκα μαζί τους και η συμβολή τους έπαιξε καθοριστικό ρόλο στην ολοκλήρωση της παρούσας πτυχιακής εργασίας κατά το ακαδημαϊκό έτος 2017-2018. Καταρχήν, θα ήθελα να εκφράσω την ειλικρινή ευγνωμοσύνη μου στον Δρ. Ποταμιάνο Γεράσιμο, Αναπληρωτή Καθηγητή του τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών στον Βόλο και κύριο επιβλέποντα της παρούσας εργασίας για τη συνεχή στήριξη, την υπομονή, τα κίνητρα και την τεράστια γνώση του. Η καθοδήγησή του με βοήθησε σε όλη την διάρκεια της έρευνας και της γραφής αυτής της διατριβής. Συγχρόνως θέλω να ευχαριστήσω τον Δρ. Αντώνιο Αργυρίου, Επίκουρο Καθηγητή του τμήματος και δεύτερο επιβλέπων της συγκεκριμένης εργασίας που είχε την καλοσύνη να διαθέσει τον πολύτιμο χρόνο του να αναγνώσει προσεκτικά την εργασία μου και να μου προσφέρει τις πολύτιμες συμβουλές και παρατηρήσεις του. Τέλος, δεν μπορώ να μην αναφερθώ στην οικογένεια μου και να εκφράσω την ευγνωμοσύνη μου για την συνεχή υποστήριξη τους, τόσο οικονομική όσο και ηθική κατά τη διάρκεια των σπουδών μου. Αλλά και ένα μεγάλο ευχαριστώ στους φίλους για την ανοχή που υπέδειξαν στο διάστημα εκπόνησης της διπλωματικής μου εργασίας καθώς και για τα φοιτητικά χρόνια που μου χάρισαν.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Συντομογραφίες

ΒΔ	**Β**άση **Δ**εδομένων
ΝΔ	**Ν**ευρωνικά **Δ**ίκτυα
ΝΔΣ	**Ν**ευρωνικά **Δ**ίκτυα **Σ**υνέλιξης
ΝΔΑ	**Ν**ευρωνικά **Δ**ίκτυα **Α**νατροφοδότησης
ΑΣΕΠ	**Α**λγόριθμος **Σ**ύγκλισης **Ε**λαχιστοποίησης **Π**αραγώγου
ΜΔΣ	**Μ**ηχανές **Δ**ιανυσματικής **Σ**τήριξης
ΕΑΜ	**Ε**πικοινωνία **Α**νθρώπου **Μ**ηχανής
ΜΕΕ	**Μ**οντέλο **Ε**νεργής **Ε**μφάνισης
ΤΔΜ	**Τ**οπικά **Δ**υαδικά **Μ**οτίβα
ΑΑΜ	**Α**λληλεπίδραση **Α**νθρώπου **Μ**ηχανής

Αφιερωμένο στην οικογένεια μου...

Κεφάλαιο 1

Εισαγωγή

Το κύριο χαρακτηριστικό που διαχωρίζει τους ανθρώπους από τις μηχανές και τους καθιστά ανώτερους ως οντότητες δεν είναι η δυνατότητα που έχουν να σκέφτονται και να επεξεργάζονται πληροφορίες, αλλά κυρίως η εγγενής ιδιότητα που τους διακατέχει να αισθάνονται, να έχουν δηλαδή συναισθήματα χαράς, λύπης, θυμού κτλ. και να δρουν σύμφωνα με αυτά. Μέσα από την εξελικτική πορεία της ανθρώπινης φύσης, η παρουσία συναισθημάτων έχει δημιουργήσει μια μορφή επικοινωνίας και αλληλεπίδρασης μεταξύ των ανθρώπων που επηρεάζει τις διαπροσωπικές σχέσεις και συνέβαλε στην υγιή συνύπαρξή τους όλα αυτά τα χρόνια. Μία απο τις εντονότερες και πιο αυθόρμητες μορφές έκφρασης αυτών των συναισθημάτων που γίνονται αντιληπτές από το ανθρώπινο μάτι είναι και οι εκφράσεις του προσώπου.

Η χρήση των μηχανών στην καθημερινότητα και η έντονη εξάρτηση του ανθρώπου από αυτές δημιουργεί την ανάγκη άμεσης κατανόησης της ανθρώπινης φύσης από την πλευρά των μηχανών. Ωστόσο, το ανθρώπινο συναίσθημα πρόκειται για μια πολυσύνθετη αντίδραση που δεν είναι τόσο εύκολο να κατηγοριοποιηθεί. Εδώ και πολλές δεκαετίες, πολλοί φιλόσοφοι, διανοητές και ψυχολόγοι έχουν εκφράσει τις απόψεις τους για το συναίσθημα. Παρόλα αυτά υπήρχαν ερωτήματα και διαφωνίες μεταξύ ερευνητών από διάφορους επιστημονικούς κλάδους και φιλοσόφους σχετικά με την ύπαρξη και τον ορισμό θεμελιωδών συναισθημάτων. Το 1994, ο Ekman [1] κατάφερε να αποτυπώσει την δικιά του θεωρία σχετικά με τον ορισμό αυτών των συναισθημάτων. Κατηγοριοποίησε λοιπόν τα συναισθήματα σε έξι βασικές κατηγορίες: τον θυμό, τον φόβο, την αηδία, την χαρά, την λύπη και την έκπληξη. Ένα μεγάλο μέρος της επιστημονικής κοινότητας ασπάστηκε αυτή την προσέγγιση και προσπάθησε να τα εντάξει στο επιστημονικό του έργο.

Η απουσία τόσα χρόνια ισχυρών υπολογιστικών συστημάτων αλλά και η συλλογή μεγάλου όγκου δεδομένων για την δημιουργία βάσεων δεδομένων αποτέλεσε εμπόδιο για την εξέλιξη τέτοιων αλγορίθμων. Τα τελευταία χρόνια ο παράλληλος προγραμματισμός αναπτύχθηκε καθώς υποστηρίζεται από σύγχρονες κάρτες γραφικών [2] ικανές να διαχειριστούν τεράστιες ποσότητες δεδομένων. Τα Νευρωνικά δίκτυα (ΝΔ) και πιο συγκεκριμένα τα νευρωνικά δίκτυα συνέλιξης (ΝΔΣ) αποτελούν ένα πολύτιμο εργαλείο στον τομέα της υπολογιστικής όρασης. Η ανάλυση του ανθρώπινου συναισθήματος μέσα απο τα φυσικά χαρακτηριστικά του προσώπου αποτελεί μια ενδιαφέρουσα πρόκληση που μπορεί να φανεί χρήσιμη σε έναν ευρύ τομέα εφαρμογών όπως τα ηλεκτρονικά παιχνίδια, η υγεία και η παρακολούθηση μέσα από κάμερα.

1.1 Σχετική έρευνα

Πολλές μελέτες έχουν διεξαχθεί στο παρελθόν με αντικείμενο την εύρεση και ανάλυση των χαρακτηριστικών του προσώπου με στόχο την αναγνώριση κάποιου συναισθήματος μέσα από αυτά. Οι τεχνικές που ακολούθησαν οι ερευνητές ποικίλουν σε μεγάλο βαθμό: ΝΔΣ, νευρωνικά δίκτυα ανατροφοδότησης (ΝΔΑ), μοντέλο ενεργής εμφάνισης (ΜΕΕ), εντοπισμός ενεργών στοιχείων, και άλλες.

Για παράδειγμα, ο Shan [3] ο οποίος χρησιμοποίησε την μέθοδο τοπικών δυαδικών μοτίβων (ΤΔΜ, local binary patterns) για την εξαγωγή των χαρακτηριστικών και μηχανές διανυσματικής στήριξης (ΜΔΣ, support vector machines) για την κατηγοριοποίηση. Τα αποτελέσματα θεωρήθηκαν αρκετά αποδοτικά για την αναγνώριση κάποιου συναισθήματος ακόμα και σε φωτογραφίες με διαφορετική ανάλυση. Επίσης, ο Dewi [4] το 2016 πρότεινε το μοντέλο ενεργής εμφάνισης (MEE, active apperance model) σε συνδυασμό με fuzzy C-means για την αναγνώριση συναισθήματος ανάμεσα σε 7 κατηγορίες. Το ΜΕΕ πρόκειται για έναν αλγόριθμο εντοπισμού προτύπων με κοινά στοιχεία που στοχεύει στην εξαγωγή χαρακτηριστικών κατά την φάση εκπαίδευσης του μοντέλου. Χρησιμοποιήθηκαν 68 σημεία για την ανάλυση του προσώπου. Για την κατηγοριοποίηση χρησιμοποιήθηκε ο fuzzy C-means και τα αποτελέσματα στην CK+ [5] βάση δεδομένων έχουν ως αποτέλεσμα 80.71% ακρίβεια.

Σε μια άλλη μελέτη ο Levi [6] συνδύασε τα ΝΔΣ με ΤΔΜ για την αναγνώριση συναισθήματος. Ο ερευνητής επικεντρώθηκε σε 2 βασικά προβλήματα: Στο περιορισμένο μέγεθος των δεδομένων καθώς και στα προβλήματα που δημιουργούνται από την διαφοροποίηση στον φωτισμό των εικόνων. Τα ΤΔΜ χρησιμοποιήθηκαν με σκοπό

την σταθεροποίηση των αποτελεσμάτων των εικόνων στις διαφοροποιήσεις του φωτισμού που στη συνέχεια αποτέλεσε είσοδο στο ΝΔΣ. Η συγκεκριμένη προσέγγιση εμφάνισε ικανοποιητικά αποτελέσματα.

Μια ακόμα ομάδα ερευνητών [7] προσέγγισε την συγκεκριμένη εργασία με την χρήση ΝΔΣ. Χρησιμοποίησε τα ΝΔΣ για την εξαγωγή των χαρακτηριστικών και όρισε ως κατηγοριοποιητή τις ΜΔΣ. Τα αποτελέσματα άγγιξαν το 98.8% στην CK+ ΒΔ και το 98.12% στην ΒΔ JAFFEE.

1.2 Συνεισφορά πτυχιακής

Σε αυτή την εργασία έχει γίνει μια προσπάθεια να αναλυθούν τα ΝΔΣ μέσα από την δοκιμασία της αναγνώρισης συναισθημάτων. Τα ΝΔΣ έχουν δείξει εξαιρετικά αποτελέσματα στον τομέα της Υπολογιστικής Όρασης. Γίνεται μια προσπάθεια να εξεταστούν σε θεωρητικό καθώς και σε πρακτικό επίπεδο μέσα απο το προγραμματιστικό κομμάτι της εργασίας όπως και οι τεχνικές οπτικής προεπεξεργασίας, οι βάσεις δεδομένων που σχετίζονται με τα φυσικά χαρακτηριστικά του προσώπου, καθώς και ο σειριακός ταξινομητής.

1.3 Οργάνωση πτυχιακής

Αυτή η πτυχιακή είναι χωρισμένη σε 6 κεφάλαια, δομημένα με σκοπό να ξετυλίξουν την πορεία μελέτης και εξέλιξης της έρευνας που πραγματοποιήθηκε για την ολοκλήρωσή της. Αφού προηγήθηκε μια σύντομη περιγραφή της εργασίας και του προβλήματος που προσπαθεί να επιλύσει, ακολουθούν τα:

- **Κεφάλαιο 2** που παρέχει πληροφορίες για τις βασικές παραμέτρους στις οποίες στηρίζονται οι βάσεις δεδομένων που σχετίζονται με θεματολογία την αναγνώρισης συναισθημάτων. Επιπλέον γίνεται αναφορά της δομής και της οργάνωσης 3 βάσεων: CK+, FER2013, MMI με εμφάνιση εικόνων και στατιστικών στοιχείων για την αναλυτική περιγραφή τους.
- **Κεφάλαιο 3** το οποίο κάνει ανάλυση των τεχνικών που χρησιμοποιήθηκαν για την επεξεργασία των εικόνων πριν την ένταξη τους στο ΝΔ.

- **Κεφάλαιο 4** παρέχει μια λεπτομερή ανάλυση της δομής και λειτουργίας των ΤΝΔ και στη συνέχεια επικεντρώνεται στα ΝΔΣ, περιγράφοντας δομικά στοιχεία που τα απαρτίζουν, την λειτουργία και τον τρόπο χρήσης τους.

- **Κεφάλαιο 5** που περιγράφει όλα τα πειράματα που εκτελέστηκαν κατά την εκπαίδευση ενός ΝΔ ικανού να προβλέπει συναισθήματα από την παρατήρηση των φυσικών χαρακτηριστικών του προσώπου. Επίσης, αναπαρίστανται τα αποτελέσματα μέσα από πίνακες και εικόνες που προσφέρουν μια καλύτερη κατανόηση τους.

- **Κεφάλαιο 6** που συνοψίζει τη Διπλωματική αυτή και περιγράφει μελλοντικές ερευνητικές κατευθύνσεις.

Κεφάλαιο 2

Βάσεις Δεδομένων (ΒΔ)

Ένα μεγάλο εμπόδιο στην ανάπτυξη του τομέα της αναγνώρισης και κατανόησης της ανθρώπινης συμπεριφοράς είναι η έλλειψη καλά οργανωμένων βάσεων δεδομένων ικανών να ακολουθήσουν τα εξέλιξη σύγχρονων αλγορίθμων νευρωνικών δικτύων (ΝΔ). Η απόδοση των ΝΔ συνδέεται άρρηκτα με την ποιότητα και την ποσότητα των δεδομένων. Ο αυθορμητισμός, η φωτεινότητα, η γωνίες, το περιβάλλον είναι μόνο κάποιες από τις παραμέτρους που μπορούν να επηρεάσουν σημαντικά της απόδοση και την ευστάθεια των μοντέλων μας. Επιπλέον, η ποικιλία φυσικών χαρακτηριστικών, η παρουσία δηλαδή δειγμάτων με διαφορετική καταγωγή, χρώμα και γενικά διάφορων στοιχείων που μπορεί να προσθέσουν την διαφορετικότητα της κάθε φυλής που μπορείς να συναντήσεις παγκοσμίως αποτελεί ένα ακόμα στοιχείο για την σύσταση μια ολοκληρωμένης βάσης δεδομένων.

Σχημα 2.1: Αναπαράσταση των 7 συναισθημάτων, δείγμα απο $CK+$ [5].

2.1 Κατηγορίες παραμετροποίησης

Οι παράμετροι που μπορούν να διαφοροποιηθούν σε μια βάση δεδομένων που απευθύνεται στην αναγνώριση φυσικών εκφράσεων του προσώπου είναι:

- **Η διέγερση** που υπόκειται το υποκείμενο για την πρόκληση του συναισθήματος. Υπάρχουν δύο τρόποι να προκαλέσεις ένα συναίσθημα. Είτε να στηθεί επιτηδευμένα το υποκείμενο σε μία έκφραση ενός συναισθήματος (posed), είτε να καταγραφεί το υποκείμενο κατά την διάρκεια αυθόρμητων μορφών έκφρασης (spontaneous) [8]. Όταν γίνεται αναφορά σε επιτηδευμένες μορφές έκφρασης συναισθημάτων αυτές δύναται να προκληθούν με πολλούς τρόπους: από προβολή ταινιών, άκουσμα ιστοριών, άκουσμα μουσικής. Για παράδειγμα στην eNTERFACE'05 EMOTION [9] βάση δεδομένων, η πρόκληση των συναισθημάτων βασίστηκε στο άκουσμα έξι διαφορετικών ιστοριών, όπου η καθεμιά προκαλούσε ένα συγκεκριμένο συναίσθημα. Στη συνέχεια, γινόταν αξιολόγηση από ειδικούς αν το συναίσθημα που προκλήθηκε είχε ξεκάθαρο χαρακτήρα. Ωστόσο, σύγχρονες έρευνες επικεντρώνονται περισσότερο στη συλλογή δεδομένων με αυθόρμητο χαρακτήρα, μια αποστολή που παρουσιάζει ιδιαίτερες δυσκολίες. Επιτυγχάνεται κυρίως μέσα από συνομιλίες μεταξύ 2 ανθρώπων. Παρ'όλα αυτά έχει επιχειρηθεί να υπάρξει αλληλεπίδραση μεταξύ μηχανής-ανθρώπου (AAM) στο έργο ELIZA [10] που αναπτύχθηκε από τον Weizenbaum.

- Παραδοσιακά η **κατηγοριοποίηση των συναισθημάτων** βασιζόταν στα 6 θεμελιώδη συναισθήματα. Η δυσκολία όμως που παρουσιάστηκε να καταγραφούν και να κατηγοριοποιηθούν όλες οι εκφράσεις σε αυτά τα συναισθήματα οδήγησε σε έρευνες που κατηγοριοποιούν τα συναισθήματα σε ένα επίπεδο δυο διαστάσεων [11]. Όπως είδαμε και παραπάνω, η καταγραφή αυθόρμητων εκφράσεων παρουσιάζει δυσκολίες, ιδιαίτερα για συναισθήματα όπως ο φόβος και η αηδία όπως και στην κατηγοριοποίηση τους, καθώς συναισθήματα όπως η βαρεμάρα ή το ενδιαφέρον για κάτι δεν μπορούν να ενταχθούν σε κάποια κατηγορία. Για αυτό τον λόγο έγινε προσέγγιση αξιολόγησης των συναισθημάτων προς ένα επίπεδο παραπάνω διαστάσεων. Στην πρώτη διάσταση εμφανίζεται η ένταση (arousal) ενός συναισθήματος και στην άλλη διάσταση η διάθεση (κατά πόσο θετική ή αρνητική) μπορεί να χαρακτηριστεί η ενέργεια που εκπέμπεται από το συναίσθημα (valence).

- **Το είδος του περιβάλλοντος**: Αν η βάση δημιουργήθηκε μέσα σε κάποιο ελεγχόμενο εργαστήριο ή σε περιβάλλον πραγματικών συνθηκών είναι η τελευταία παράμετρος που επηρεάζει. Συνήθως οι βάσεις που δημιουργήθηκαν σε εργαστήρια χρησιμοποιούν στατικό φοντο και ελεγχόμενο φωτισμό, ενώ σε πραγματικές συνθήκες παρουσιάζονται περισσότερα χρώματα στο φόντο.

2.2 ΒΔ για την αναγνώριση συναισθημάτων από εκφράσεις του προσώπου

Η επιστημονική κοινότητα έχει δημιουργήσει τα τελευταία χρόνια μια λίστα από βάσεις δεδομένων που απευθύνονται στην αναγνώριση συναισθήματος μέσα από τις εκφράσεις των φυσικών χαρακτηριστικών του προσώπου όπως οι: Cohn Kanade και Cohn Kanade+ [5], FER2013 [12] καθώς επίσης και MMI [13] και η βάση του Toronto [14].

2.2.1 Extended Cohn-Kanade Dataset (CK+)

Στα πλαίσια της συγκεκριμένης πτυχιακής χρησιμοποιείται ως βάση δεδομένων The Extended Cohn-Kanade database που δημιουργήθηκε με σκοπό τον αυτόματο εντοπισμό εκφράσεων του προσώπου. Πρόκειται για μία επέκταση της αρχικής έκδοσης Cohn-Kanade Dataset (CK+) [15] που δημοσιεύτηκε για πρώτη φορά το 2000. Η συγκεκριμένη βάση δεδομένων είναι διαθέσιμη για ερευνητικούς σκοπούς σε 2 εκδόσεις ουσιαστικά και μία τρίτη έκδοση βρίσκεται αυτή την στιγμή σε προετοιμασία. Περιλαμβάνει 486 συστοιχίες εικόνων που περιλαμβάνουν εκφράσεις προσώπου σε posed μορφή από 97 υποκείμενα στην πρώτη έκδοση και επεκτείνεται στην δεύτερη έκδοση καταλήγοντας με 593 συστοιχίες εικόνων που περιλαμβάνουν εκφράσεις προσώπου σε posed μορφή ανάμεσα σε 123 υποκείμενα στην δεύτερη έκδοση (CK+). Κάθε έκφραση περιλαμβάνει εικόνες που ξεκινούν από ένα ουδέτερο πλάνο και καταλήγουν στην κορύφωση της έκφρασης του συναισθήματος μέσα από μια σειρά εικόνων. Παρακάτω στους πίνακες 2.1 ακολουθούν κάποια στατιστικά στοιχεία σχετικά με τα δείγματα κάθε έκφρασης που περιλαμβάνονται στην βάση δεδομένων.

Σχημα 2.2: Πλάνα από την βάση δεδομένων $CK+$ [5] από την στιγμή της ουδέτερης στάσης μέχρι το σημείο κορύφωσης της έκφρασης του συναισθήματος.

Τα δείγματα της που χρησιμοποιήθηκαν στην βάση δεδομένων CK+ βρίσκονταν ηλικιακά μεταξύ 18-50 έτη. Το 69% των δειγμάτων αποτελείται από γυναίκες ενώ το υπόλοιπο 21% από άντρες. Τέλος με βάση την καταγωγή των δειγμάτων το 81% άνηκε σε Euro-Americans, το 13% σε Afro-Americans και το υπόλοιπο σε κάποια άλλη ομάδα. Η συγκεκριμένη βάση δεδομένων ταξινομεί κάθε έκφραση το προσώπου ανάμεσα σε 7 θεμελιώδεις κατηγορίες: τον θυμό, την απέχθεια, τον φόβο, την χαρά, την αηδία, την λύπη και την έκπληξη

Συναίσθημα	**Αριθμός Εκφράσεων**
Θυμός	45
Απέχθεια	18
Αηδία	59
Φόβος	25
Χαρά	69
Λύπη	28
Έκπληξη	83

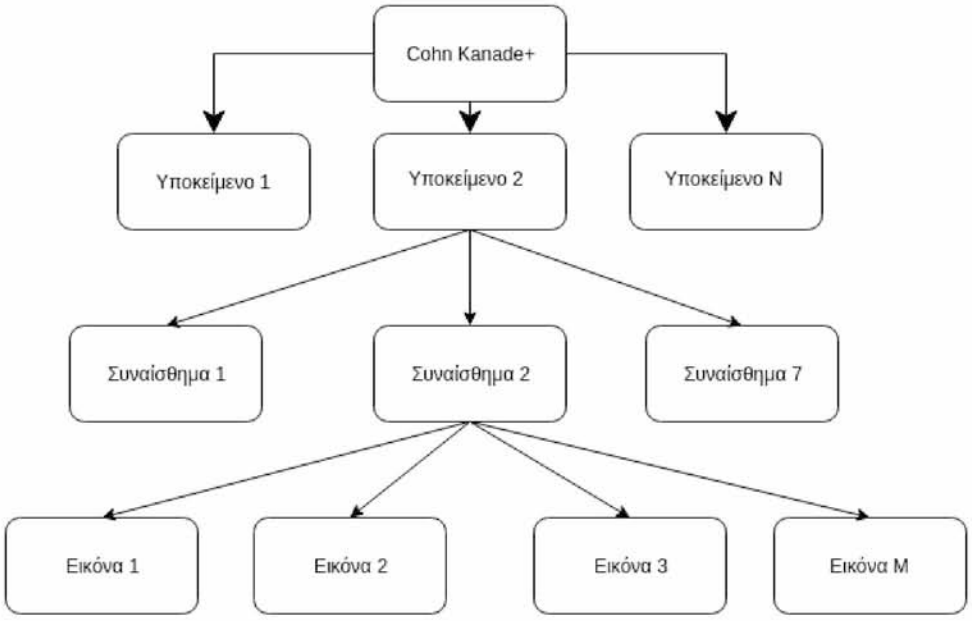
ΠΙΝΑΚΑΣ 2.1: Συχνότητα έκφρασης συναισθημάτων στην βάση CK+.

Η τρίτη έκδοση ειναι προγραμματισμένη να εκδοθεί σύντομα. Η αρχική μορφή της βάσης δεδομένων περιλαμβάνει εικόνες με τα υποκείμενα να ποζάρουν ευθυγραμμισμένα με την κάμερα. Στην ανανεωμένη έκδοση περιλαμβάνεται και υλικό με τα υποκείμενα να έχουν κλήση 30 μοίρες.

2.2.1.1 Δομή φακέλων της βάσης

Η βάση δεδομένων CK+ είναι δομημένη σε φακέλους που έχουν την παρακάτω μορφή: Κάθε υποκείμενο τοποθετήθηκε σε έναν ξεχωριστό φάκελο και μέσα σε αυτούς τους φακέλους περιλαμβάνονται υποφάκελοι που ο καθένας κατηγοριοποιείται σε ένα συγκεκριμένο συναίσθημα.

Στον τελευταίο υποφάκελο βρίσκονται οι εικόνες του υποκειμένου που ξεκινάνε από μια εικόνα που το απεικονίζει σε ουδέτερη στάση μέχρι το τελευταίο πλάνο στο οποίο απεικονίζεται στην κορύφωση της έκφρασης του επιλεγμένου συναισθήματος.



Σχημα 2.3: Δομή φακέλων στην βάση δεδομένων CK+.

Ακριβώς την ίδια δομή έχουν και οι φάκελοι με τις ετικέτες κάθε συναισθήματος. Όλες οι ετικέτες είναι αποθηκευμένες σε αρχείο μορφής text και αναπαριστούν το συναίσθημα με έναν αριθμό από το 1 μέχρι το 7 που δηλώνει ένα συγκεκριμένο συναίσθημα.

2.2.2 FER2013 Dataset

Η βάση δεδομένων FER2013 δημιουργήθηκε από τους Pierre Luc Carrier, Aaron Courville και αποτελεί μέρος ενός μεγαλύτερου έργου. Για την κατασκευή της χρησιμοποιήθηκε το Google API που εξειδικεύεται στην αναζήτηση εικόνων. Η αναζήτηση επικεντρώθηκε σε εικόνες από πρόσωπα που ταιριάζουν σε ένα γκρουπ λέξεων που περιγράφουν συναισθήματα όπως: ευτυχισμένος, εξοργισμένος κτλ. Αυτά τα συναισθήματα συνδυάστηκαν με λέξεις που σχετίζονται με την ηλικία, το φύλο, την εθνικότητα και 600 ακόμα κατηγορίες. Στη συνέχεια, οι πρώτες 1000 λέξεις από κάθε κατηγορία προχώρησαν στο επόμενο στάδιο το οποίο περιελάμβανε: τον εντοπισμό της περιοχής του προσώπου με την χρήση εργαλείων το OpenCV. Παρακάτω, εντοπίστηκαν οι εικόνες που περιείχαν λανθασμένες ετικέτες καθώς και διπλότυπες εικόνες. Στο τέλος, οι εγκεκριμένες εικόνες έλαβαν τις διαστάσεις 48×48 pixels και μετατράπηκαν σε greyscale.

Ο Mehdi Mirza και ο Ian Goodfellow χώρισαν τις ετικέτες από τις εικόνες σε 7 κατηγορίες σύμφωνα με την βάση δεδομένων του Toronto. Το τελικό μέγεθος της βάσης δεδομένων αποτελείται από 35887 εικόνες. Αν και αρχικά υπήρχαν επιφυλάξεις για τον τρόπο που συλλέχθηκαν οι εικόνες και για την κατηγοριοποίηση τους ο Ian Goodfellow εντόπισε πως δεν επηρεάζει σημαντικά την εκπαίδευση ενός ΝΔ.

Ο αριθμός των εικόνων σε κάθε κατηγορία εμφανίζεται παρακάτω:

Συναίσθημα	**Εικόνες**
Θυμός	4953
Αηδία	547
Φόβος	512
Χαρά	8989
Λύπη	6077
Έκπληξη	4002
Ουδέτερο	6198

Πινακας 2.2: Κατηγορίες εικόνων βάση συναισθήματος FER2013.

2.2.3 MMI Dataset

Η βάση δεδομένων MMI [13] δημιουργήθηκε αρχικά το 2002 απο τους Maja Pantic, Michel Valstar, Ioannis Patras εχοντας ως στόχο την αξιολόγηση αλγορίθμων που σχετίζονται με την αναγνώριση εκφράσεων του προσώπου και κατ' επέκταση την συνεισφορά στην κοινότητα οπτικών μέσων ενός μεγάλου όγκου δεδομένων οπτικού υλικού. Η συγκεκριμένη βάση δεδομένων περιλαμβάνει 2900 βίντεο, υψηλής ποιότητας εικόνες από 79 υποκείμενα. Η βάση δεδομένων MMI προσεγγίζει το θέμα της αναγνώρισης συναισθημάτων όχι μόνο μέσα από τον κατηγοριοποίηση τους σε 7 θεμελιώδεις κατηγορίες, αλλά περιλαμβάνει επιπλέον πληροφορίες για την ενεργοποίηση συγκεκριμένων μυικών ομάδων του προσώπου και άλλους δείκτες περιγραφής. Η συγκεκριμένη βάση δεδομένων διαθέτει στατικές εικόνες καθως και συστοιχίες εικόνων από τα πρόσωπα των αντικειμένων σε ανφας και σε προφίλ. Όλες οι εικόνες απεικονίζονται απο 24-bit, 720×576 ανάλυση. Όλα τα βίντεο έχουν καταγραφεί με ένα ρυθμό 24 πλάνων το δευτερόλεπτο χρησιμοποιώντας μια PAL κάμερα. Το 25% των δειγμάτων έχουν καταγραφεί με φυσικό φωτισμό και μια ποικιλία από φόντα, ενώ τα υπόλοιπα δείγματα έχουν ως φόντο ένα μπλε πλαίσιο και υψήλης έντασης φωτισμό με αντανακλαστικές ομπρέλες.

Η βάση δεδομένων MMI έχει αναπτυχθεί επιπλέον ως μια web εφαρμογή άμεσης χρήσης διαθέσιμη σε συγκεκριμένα μέλη που έχουν την αντίστοιχη εξουσιοδότηση.

Η αναζήτηση μέσα σε αυτή τήν βάση από την εφαρμογή μπορεί να πραγματοποιηθεί με ευέλικτους τρόπους σύμφωνα με τα κριτήρια της αναζήτησης: οπτική γωνία, ενεργά σημεία, φύλο, ηλικία κτλ. Η βάση MMI επιτρέπει την ενοποίηση με άλλες βάσεις δεδομένων με εύκολο και φιλικό τρόπο προς τους χρήστες.

ΒΔ	**Περιβάλλον**	**Ηλικία**	**Φωτισμός**	**Υποκείμενα**
AFEW [16]	ΚΣΠ	1-70	ΚΣΦ	330
Belfast [17]	*TV* & Εργαστήριο	;	Ε	100
CK [5]	Εργαστήριο	18-50	Ε	97
CK+ [15]	Εργαστήριο	18-50	Ε	123
F.TUM	Εργαστήριο	;	Ε	18
GEMEP [18]	Εργαστήριο	;	Ε	10
M − PIE [19]	Εργαστήριο	27	Ε	337
MMI	Εργαστήριο	19-62	Ε	29
RU − FACS [20]	Εργαστήριο	18-30	Ε	100
Semaine [8]	Εργαστήριο	;	Ε	75
UT − Dallas [21]	Εργαστήριο	18-25	Ε	284
EMOTIC [22]	ΚΣΠ	;	ΚΣΦ	;
VAM [23]	ΚΣΠ	;	Ε	20

ΠΙΝΑΚΑΣ 2.3: Περιγραφή βάσεων δεδομένων που χρησιμοποιούνται στην αναγνώριση εκφράσεων του προσώπου, ΚΣΠ = Κοντά στο πραγματικό περιβάλλον, ΚΣΦ = Κοντά στον φυσικό φωτισμό, Ε = ελεγχόμενο.

Κεφάλαιο 3

Τεχνικές Οπτικής Προεπεξεργασίας

Είναι γεγονός πως τα ΝΔ και συγκεκριμένα τα ΝΔΣ έχουν αυξημένα ποσοστά ακρίβειας σε προβλήματα κατηγοριοποίησης οπτικών δεδομένων, παρόλα αυτά παρατηρείται πως ορισμένες παράμετροι επηρεάζουν δραστικά την αξιοπιστία αυτών των αλγορίθμων. Μία από αυτές τις παραμέτρους είναι ο φωτισμός της εικόνας. Για αυτό τον λόγο χρησιμοποιούνται διάφορες τεχνικές που δύνανται να διορθώσουν τις συνθήκες φωτισμού. Πέρα από τους αλγορίθμους που σχετίζονται με αυτή την παράμετρο, υπάρχουν και οι αλγόριθμοι προεπεξεργασίας που στοχεύουν στην διαχείριση των δεδομένων και την ομαλοποίηση των τιμών. Στα πλαίσια αυτής της εργασίας εφαρμόστηκαν τέτοιες τεχνικές προεπεξεργασίας όπως θα δούμε και παρακάτω για την επίτευξη μέγιστης ακρίβειας.

3.1 Μέθοδοι προεπεξεργασίας για την ρύθμιση του φωτισμού

Κάποιες από αυτές τις παραμέτρους που επιδρούν στα αποτελέσματα των αλγορίθμων είναι ο φωτισμός κατά την λήψη των εικόνων, η ποιότητα των εικόνων και ο όγκος των δεδομένων. Μάλιστα, ενδέχεται να παρατηρηθούν μεγαλύτερες αποκλίσεις στα αποτελέσματα εικόνων που απεικονίζουν μία συγκεκριμένη έκφραση προσώπου σε διαφορετικές συνθήκες φωτισμού σε σύγκριση με εικόνες που απεικονίζουν διαφορετικές εκφράσεις του προσώπου. Έτσι κρίνεται αναγκαίο να εφαρμοστούν κάποιες

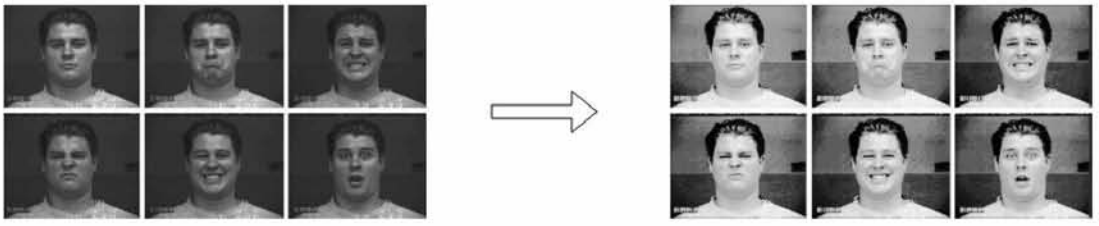
τεχνικές επεξεργασίας των εικόνων πριν εισαχθούν στο ΝΔΣ για να αντιμετωπιστούν τέτοιου είδους προβλήματα. Τέτοιοι αλγόριθμοι που επικεντρώνονται σε θέματα με τον φωτισμό των εικόνων είναι ο αλγόριθμος διόρθωσης της έντασης της τιμής gamma, ο μετασχηματισμός λογαρίθμου, η εξισορρόπηση του ιστογράμματος και ο διακριτός μετασχηματισμός του συνημίτονου. Στη συγκεκριμένη εργασία υλοποιήθηκε και εφαρμόστηκε ο αλγόριθμος εξισορρόπησης ιστογράμματος.

3.1.1 Εξισορρόπηση ιστογράμματος

Η παρουσία των ασπρόμαυρων εικόνων στην εργασία επιτάσσει μια σύντομη περιγραφή για το πως δουλεύουν. Κάθε pixel της εικόνας αναπαριστάται από μια τιμή που εκφράζει την ένταση του χρώματος. Οι τιμές κυμαίνονται απο [0, 255]. Αποτελούνται αποκλειστικά από αποχρώσεις του γκρι, όπου το μαύρο ισοδυναμεί με την μικρότερη ένταση και το άσπρο την εντονότερη.

Η εξισορρόπηση ιστογράμματος (histogram equalization) είναι μια τεχνική που έχει ως στόχο να ρυθμίσει την ένταση των pixels σε μία εικόνα. Ανακατανέμεται το αρχικό ιστόγραμμα σε ολόκληρη την κλίμακα διακριτών τιμών της εικόνας με στόχο να ενισχυθεί η αντίθεση [24].

Το ιστόγραμμα μιας εικόνας μπορεί να αναπαρασταθεί απεικονίζοντας σε άξονες την ένταση των pixels στον ένα άξονα και την συχνότητα στον άλλον ή την πιθανότητα εμφάνισης.



Σχημα 3.1: Πριν και μετά από την εξισορρόπηση του ιστογράμματος. Δείγμα εικόνων από την ΒΔ CK+ [15]

3.2 Μέθοδοι προεπεξεργασίας για την σωστή διαχείριση δεδομένων

3.2.1 Κανονική ομαλοποίηση

Η κανονική ομαλοποίηση ή normalization είναι μια τεχνική που εκτελείται με στόχο να ρυθμίσει όλα τα δεδομένα έτσι ώστε να βρίσκονται στην ίδια κλίμακα τιμών. Η βασική ιδέα είναι πως τα δεδομένα που έχουμε στην διάθεση μας είναι πιθανό να περιλαμβάνουν τιμές που αποκλίνουν σε μεγάλο βαθμό επηρεάζοντας με αυτόν τον τρόπο τα αποτελέσματα του ΝΔ. Η ομαλοποίηση μπορεί να επιτευχθεί στις εικόνες αφαιρώντας από την τιμή του κάθε pixel την μέση τιμή και στη συνέχεια να διαιρεθεί το αποτέλεσμα με την τυπική απόκλιση. Η ομαλοποίηση συμβάλλει επίσης στην γρηγορότερη εκπαίδευση του δικτύου

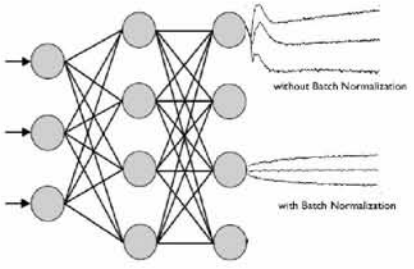
$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \quad \text{(Μέση τιμή)}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu)^2 \quad \text{(Διακύμανση τιμών)}$$

$$\hat{x} = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad \text{(Ομαλοποίηση τιμών)}$$

3.2.2 Ομαλοποίηση σε πακέτα

Η ομαλοποίηση σε πακέτα ή batch normalization, πραγματοποιείται για να μειώσει την απόκλιση της κατανομής των δεδομένων εκπαίδευσης σε σχέση με τα δεδομένα ελέγχου [25] ή αλλιώς covariate shift. Με αυτό τον τρόπο επιταχύνεται η διαδικασία εκπαίδευσης του ΝΔ καθώς επίσης βοηθάει και στην αντιμετώπιση του φαινομένου της υπερπροσαρμογής προσθέτοντας θόρυβο στο ΝΔ και οδηγώντας κάθε επίπεδο να λειτουργεί σχετικά ανεξάρτητα από τα υπόλοιπα επίπεδα του δικτύου.



Σχημα 3.2: Τα δεδομένα ύστερα από την χρήση ομαλοποίησης κατά πακέτα (σχήμα από [25]).

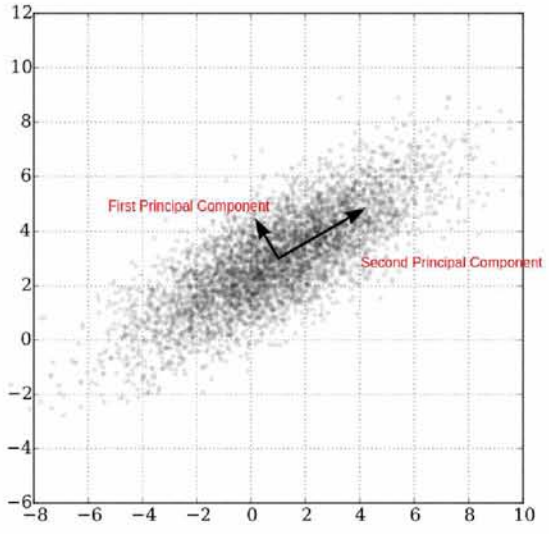
Η ομαλοποίηση σε ομάδες αποτελείται από 2 αλγορίθμους. Ο πρώτος αλγόριθμος σχετίζεται με την μετατροπή των δεδομένων εισόδου x στις νέες τιμές y που έχουν μετακινηθεί και έχουν υποστεί ομαλοποίηση. Ο δεύτερος αλγόριθμος σχετίζεται με την εκπαίδευση του ομαλοποιημένου κατά ομάδες ΝΔ. Κατά την υλοποίηση του πρώτου αλγορίθμου για την μετακίνηση καθώς και την ομαλοποίηση των δεδομένων εισάγονται 2 ακόμα παράμετροι γ, β που συμμετέχουν στον αλγόριθμο οπισθοδρόμησης και οι τιμές τους είναι μεταβλητές.

$$y = \gamma * \hat{x} + \beta \quad \text{(Κλιμάκωση και μετακίνηση τιμών)}$$

3.2.3 Ανάλυση κυρίων συνιστωσών (PCA)

Το PCA είναι μια τεχνική που αφορά την περιγραφή και αναπαράσταση πολυδιάστατων δεδομένων χρησιμοποιώντας την μέγιστη διακύμανση [27]. Είναι ένα χρήσιμο εργαλείο που χρησιμοποιείται στην μείωση των διαστάσεων για την καλύτερη διαχείριση της πληροφορίας καθώς και στην μείωση του όγκου δεδομένων. Τα βήματα που ακολουθεί ο αλγόριθμος είναι:

- Ομαλοποίηση των δεδομένων
- Υπολογισμός του πίνακα διασποράς
- Υπολογισμός ιδιοτιμών και ιδιοδιανυσμάτων του πίνακα διασποράς
- Υπολογισμός αποτελεσμάτων



Σχημα 3.3: Ανάλυση δεδομένων στους βασικούς άξονες (σχήμα από [26]).

Σε δεδομένα που έχουν ως κέντρο το σημείο (0, 0) υπολογίζουμε τον πίνακα διασποράς. Από την διαγώνιο αυτού του πίνακα υπολογίζουμε τις διακυμάνσεις. Εκτελούμε παραγοντοποίηση ιδιόμορφων τιμών των δεδομένων και αφαιρούμε τις διαστάσεις με

τις μικρότερες τιμές διακύμανσης. Με αυτόν τον τρόπο διατηρούμε τις διαστάσεις των δεδομένων που έχουν κυρίαρχο ρόλο.

3.2.4 Προσαύξηση δεδομένων

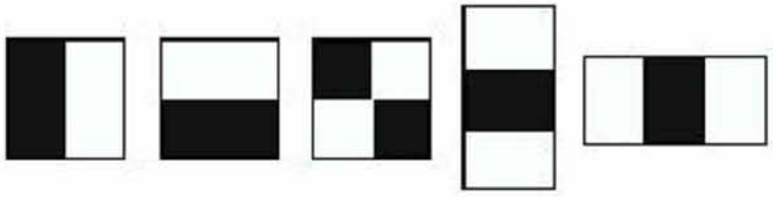
Η αποτελεσματικότητα των περισσότερων προβλημάτων υπολογιστικής όρασης συνδέεται άμεσα με το μέγεθος της ΒΔ. Ένα από τα εμπόδια που παρουσιάζονται σε πολλά προβλήματα υπολογιστικής όρασης είναι η έλλειψη μεγάλου όγκου δεδομένων που διαθέτουν πληροφορία σε ποια κατηγορία ανήκουν, καθώς αποτελεί συνήθως μια εργασία που πραγματοποιείται χειροκίνητα και έχει αρκετούς κανονισμούς. Για αυτό το λόγο η προσαύξηση των δεδομένων ή data augmentation αποτελεί ένα πολύ δυνατό εργαλείο [28] που χρησιμοποιείται συχνά στα ΝΔΣ ξεπερνώντας το εμπόδιο που αναφέρθηκε παραπάνω. Η προσαύξηση των δεδομένων είναι μια τεχνική που χρησιμοποιείται για να αυξήσει τον όγκο δεδομένων βασισμένη σε μορφοποιήσεις που συμβαίνουν στην ήδη υπάρχουσα πληροφορία. Αντιμετωπίζει αποτελεσματικά συχνά προβλήματα των ΝΔΣ όπως η υπερπροσαρμογή και προσδίδει στο δίκτυο σταθερότητα και αμεταβλητότητα των αποτελεσμάτων ως προς τις τροποποιήσεις και τους μετασχηματισμούς των αναπαραστάσεων που σχετίζονται είτε με τον φωτισμό, είτε με την θέση των αντικειμένων. Ορισμένες από τις ενέργειες που παρέχει η συνάρτηση ImageDataGenerator από την βιβλιοθήκη keras [29] είναι :

- οριζόντια περιστροφή
- κάθετη περιστροφή
- μεγέθυνση
- αλλαγή κλίμακας
- ομαλοποίηση
- μετακίνηση της εικόνας προς όλες τις κατευθνσεις
- αποκοπή σημείων της εικόνας
- zca whitening

3.3 Σειριακός ταξινομητής Haar

Ο Viola και Jones [30] εισήγαγαν στον χώρο της υπολογιστικής όρασης μία μέθοδο που βοηθάει στην ανίχνευση των φυσικών χαρακτηριστικών του προσώπου μέσω του σειριακού ταξινομητή Haar. Όλη η διαδικασία μπορεί να χωριστεί σε 3 μέρη:

- Το πρώτο μέρος περιλαμβάνει την δημιουργία των χαρακτηριστικών τύπου Haar που αποτελούν τον βασικό πυρήνα του ταξινομητή. Αυτά τα χαρακτηριστικά στηρίζουν την λειτουργία τους στην απότομη αλλαγή έντασης γειτονικών ομάδων από pixels. Αυτές οι γειτονικές ομάδες έχουν ορθογώνιο σχήμα. Όπως βλέπουμε και στην εικόνα παρακάτω υπάρχουν 3 είδη ορθογωνίων και ο υπολογισμός του αποτελέσματος ισοδυναμεί με το άθροισμα της έντασης των pixels που βρίσκονται στην λευκή περιοχή αφαιρούμενο από το αντίστοιχο άθροισμα της μαύρης περιοχής.



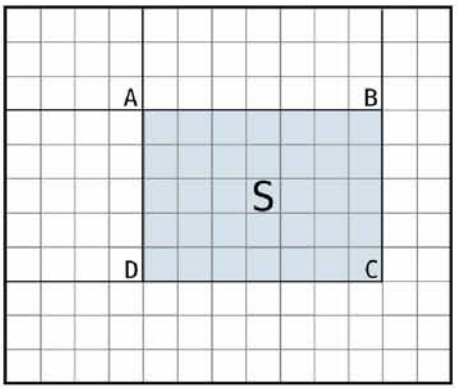
Σχημα 3.4: Εικόνες από χαρακτηριστικά τύπου Haar (σχήμα από [31]).

Τα χαρακτηριστικά αυτά υπολογίζονται με την βοήθεια των εικόνων ολοκλήρωσης τα οποία είναι μια ενδιάμεση αναπαράσταση της αρχικής εικόνας. Πρόκειται για εικόνες που το κάθε στοιχείο τους περιέχει σαν τιμή το άθροισμα των τιμών των επάνω και αριστερά στοιχείων προστιθέμενη στην δικιά του τιμή, δηλαδή:

$$I(x,y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x',y')$$

Με αυτό τον τρόπο είναι εφικτός ο υπολογισμός του αθροίσματος κάθε περιοχής ανεξάρτητα από το μέγεθος του με την χρήση μόνο τεσσάρων τιμών. Ο τύπος για τον υπολογισμό της τιμής κάθε στοιχείου είναι για τα σημεία: A(x_0, y_0), B(x_1, y_0), C(x_0, y_1), D(x_1, y_1)

$$\sum_{\substack{x_0 \leq x \leq x_1 \\ y_0 \leq y \leq y_1}} i(x,y) = I(D) + I(A) - I(B) - I(C)$$



Σχημα 3.5: Υπολογισμός τιμών σε εικόνες ολοκλήρωσης σε ορισμένη περιοχή (A, B, C, D) (σχήμα από [31]).

- Η επιλογή των κατάλληλων χαρακτηριστικών αποτελεί το δεύτερο μέρος. Αυτό επιτυγχάνεται με τον αλγόριθμο Adaboost που είναι υπεύθυνος να διαλέξει τα χαρακτηριστικά και να εκπαιδεύσει τον ταξινομητή.

- Το τελευταίο βήμα αποτελείται από την συνένωση πολλών ταξινομητών για την σύνθεση ενός ενιαίου σύνθετου ταξινομητή.

Κεφάλαιο 4

Βαθιά Νευρωνικά Δίκτυα και Υπολογιστική Όραση

4.1 Ιστορική αναδρομή

Η μελέτη του οπτικού φλοιού συνδέεται στενά με την εξέλιξη των ΝΔΣ. Το 1968 οι Hubel και Wiesel [32] παρουσίασαν μία μελέτη τους που επικεντρωνόταν στο πεδίο υποδοχής του οπτικού φλοιού των μαϊμούδων. Η μελέτη αυτή σχετίστηκε με τα νευρωνικά δίκτυα εξαιτίας της αρχιτεκτονικής του οπτικού φλοιού των μαϊμούδων και τον τρόπο που οι νευρώνες τους συνδέονταν μεταξύ τους. Οι νευρώνες αυτοί διακρίνονται σε 2 κατηγορίες: στους απλούς και στους σύνθετους. Οι απλοί νευρώνες δύνανται να εντοπίζουν γωνίες σε σχήματα, ενώ οι σύνθετοι αναγνωρίζουν πιο σύνθετα αντικείμενα και λεπτομέρειες που δεν επηρεάζονται από μετακινήσεις στον χώρο. Με αυτό τον τρόπο ο συνδυασμός των νευρώνων επιτρέπει την οπτικοποίηση μια περιοχής χρησιμοποιώντας την συσχέτιση σχημάτων και αντικειμένων.

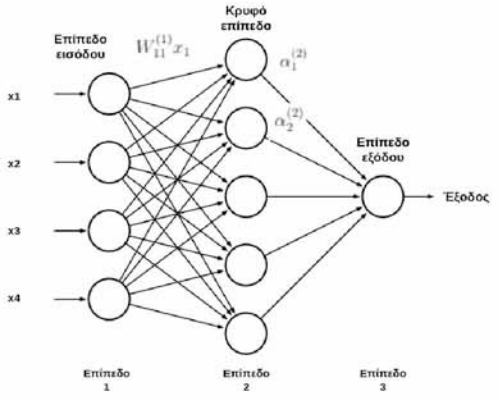
Μια από τις πρώτες υλοποιήσεις εμπνευσμένες από την ιδέα των Hubel και Wiesel είναι το Neocognitron [33]. Πρόκειται για ένα ιεραρχικό, πολυεπίπεδο τεχνητό νευρωνικό δίκτυο που προτάθηκε από το Kunihiko Fukushima στη δεκαετία του 1980. Το πρώτο επίπεδο αποτελείται από μονάδες που αναπαριστούν τους απλούς νευρώνες ενώ οι μονάδες του επόμενου επιπέδου αναπαριστούν τους σύνθετους. Το μεγαλύτερο επίτευγμα αυτού τους αλγορίθμου είναι η τοπική αμεταβλητότητα που πέτυχε, παρ'ολα αυτά διέθετε αρκετά αρνητικά. Το μεγαλύτερο ελάττωμά σε αυτό τον αλγόριθμο είναι πως δεν διέθετε κάποιο μηχανισμό για την βελτίωση της διαδικασίας εκπαίδευσης του

δικτύου ή δυνατότητα ρύθμισης κάποιων παραμέτρων. Όλα αυτά μέχρι την στιγμή που εμφανίστηκε ο μαθηματικός Seppo Linnainmaa που δημιούργησε τον αλγόριθμο οπισθοδιάδοσης ή backpropagation στη σύγχρονη του μορφή το 1970. Εως το 1985 ο αλγόριθμος αυτός χρησιμοποιήθηκε ελάχιστα σε εφαρμογές μέχρι την στιγμή που οι Rumelhart, Hinton και Williams εισήγαγαν την χρήση του αλγορίθμου οπισθοδιάδοσης στα Νευρωνικά Δίκτυα και επέφεραν την ραγδαία ανάπτυξη τους σε συνδυασμό με την ανάπτυξη της υπολογιστικής ισχύς. Ο αλγόριθμος περιγράφεται στο παρακάτω κεφάλαιο λεπτομερώς.

Στη συνέχεια εμφανίστηκε ο LeCun ο οποίος θεωρήθηκε πρωτοπόρος στην έρευνα των ΝΔΣ. Η πρώτη ουσιαστική εφαρμογή του αλγορίθμου της οπισθοδιάδοσης αποτέλεσε δικιά του δημιουργία και είχε ως στόχο την κατηγοριοποίηση εφαρμοσμένη σε χειρόγραφα ψηφία (MNIST). Αυτή η εφαρμογή αποτέλεσε την μεγαλύτερη επιτυχία των ΝΔΣ μέχρι εκείνη την στιγμή καθώς κατάφερε να ψηφιοποιήσει επιτυχώς μεγάλο αρχείο δεδομένων και αποτέλεσε έμπνευση για μελλοντικές δημιουργίες.

4.2 Τεχνικά νευρωνικά δίκτυα

Τα τεχνικά νευρωνικά δίκτυα χαρακτηρίζονται από ένα σύνολο διασυνδεδεμένων νευρώνων τα οποία μπορούν να χωριστούν σε 3 βασικά επίπεδα λειτουργίας. Τους νευρώνες που βρίσκονται στο επίπεδο εισόδου (l_1), στο κρυφό επίπεδο (l_2), το οποίο μπορεί να αποτελείται από ένα ή και περισσότερα επίπεδα και στο επίπεδο εξόδου (l_3). Η έξοδος των νευρώνων του επιπέδου (l_1) αποτελούν είσοδο για τους νευρώνες επιπέδου (l_2) και αντίστοιχα η έξοδος των νευρώνων κάθε επιπέδου αποτελούν είσοδο για τους νευρώνες του επόμενου επιπέδου.



Σχημα 4.1: Αναπαράσταση ενός τεχνικού νευρωνικού δικτύου.

Το ΝΔ χαρακτηρίζεται από κάποιες παραμέτρους $W_{i,j}^{(l)}$ όπου περιγράφει το βάρος που σχετίζεται με την σύνδεση του i-οστού νευρώνα του επιπέδου l με τον j-οστό νευρώνα του επιπέδου $l+1$ και μία παράμετρο $b_l^{(i)}$ που αποτελεί μια σταθερά bias του επιπέδου l του νευρώνα i και βοηθάει στην καλύτερη εκπαίδευση του ΝΔ. Οι ενεργοποιήσεις κάθε νευρώνα αναπαρίστανται από την τιμή α_i^l του νευρώνα i στο επίπεδο l. Στο επίπεδο $l = 1$, οι τιμές $\alpha_i^1 = x_i$, όπου x_i το i-οστό διάνυσμα εισόδου. Στο επίπεδο εξόδου υπολογίζεται η $h_{W,b}(x)$ συνάρτηση υπόθεσης η οποία υπολογίζει το άθροισμα αφού περάσει μέσα από μια συνάρτηση ενεργοποίησης και εξάγει έναν πραγματικό αριθμό. Η συγκεκριμένη αρχιτεκτονική αποτελεί ενδεικτική απλουστευμένη μορφή για την καλύτερη κατανόηση της λειτουργίας του ΝΔ. Το δίκτυο μας δύναται να είτε να έχει διαφορετικό τρόπο διασύνδεσης μεταξύ των νευρώνων, είτε να έχει άλλη μορφολογία. Ανάλογα με το πρόβλημα που αντιμετωπίζεται κάθε φορά προσαρμόζεται και η αρχιτεκτονική του δικτύου. Έτσι είναι πιθανό σε ένα πρόβλημα πολλαπλής κατηγοριοποίησης να εμφανίζονται στο επίπεδο εξόδου παραπάνω νευρώνες.

$$\alpha_1^{(2)} = f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)})$$
$$\alpha_2^{(2)} = f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)})$$
$$\alpha_3^{(2)} = f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)})$$
$$h_{W,b}(x) = \alpha_1^{(3)} = f(W_{11}^{(2)}\alpha_1^2 + W_{12}^{(2)}\alpha_2^2 + W_{13}^{(2)}\alpha_3^2 + b_1^{(2)})$$

Για απλούστευση το συνολικό άθροισμα μπορεί να πάρει την μορφή

$$z_i^{(l+1)} = \sum_{j=1}^{n}(W_{i,j}^{(l)}x_j + b_j^{(l)})$$

$$\alpha_i^{(l)} = f(z_i^{(l)})$$

Πρόκειται για μία προς τα εμπρός διάδοση σήματος (forward propagation) καθώς όπως παρατηρούμε, οι νευρώνες κάθε επιπέδου είναι αποτέλεσμα υπολογισμών που προέρχονται από το προηγούμενο επίπεδο σε ένα πλήρως διασυνδεδεμένο δίκτυο.

4.3 Αλγόριθμος οπισθοδιάδοσης

Ο αλγόριθμος της οπισθοδιάδοσης αποτελεί τον κεντρικό πυρήνα πάνω στον οποίο βασίζεται η λειτουργία όλων των αλγορίθμων των ΝΔ.

Διαισθητικά ο αλγόριθμος της οπισθοδιάδοσης υπολογίζει την κλίση του λάθους συναρτήσει τα βάρη των νευρώνων. Η κλίση αλλάζει κάθε φορά που μεταβάλλονται οι τιμές των βαρών. Στόχος του αλγορίθμου είναι να βρει εκείνες τις τιμές των βαρών που ελαχιστοποιούν την συνάρτηση κόστους στο δίκτυο. Όταν χρησιμοποιείται ο αλγόριθμος οπισθοδιάδοσης σε συνδυασμό με κάποιον αλγόριθμο βελτιστοποίησης όπως αυτός της σύγκλισης με ελάττωση της παραγώγου (ΑΣΕΠ) ή gradient descent, το σύστημα είναι έτοιμο να ρυθμίσει αυτές τις παραμέτρους αυτόματα. Ο ΑΣΕΠ πρόκειται για έναν αλγόριθμο πρώτου βαθμού που βρίσκει το τοπικό ελάχιστο της συνάρτησης κόστους κάνοντας συμμετρικά βήματα που εξαρτώνται από τον ρυθμό εκμάθησης μειώνοντας την κλίση ανάλογα.

Έτσι αν υποθέσουμε πως διαθέτουμε δεδομένα εκπαίδευσης $(x^{(1)}, y^{(1)}), ..., (x^{(m)}, y^{(m)})$ σε m δείγματα εκπαίδευσης. Η συνάρτηση κόστους για ένα δείγμα x, y ορίζεται ως:

$$J(W, b : x, y) = \frac{1}{2}\left\|h_{W,b}(x) - y\right\|^2$$

Στη συνέχεια θα πρέπει να εφαρμοστεί ένας αλγόριθμος βελτιστοποίησης για την εύρεση των παραμέτρων W, b που πρόκειται να ελαχιστοποιήσουν τη συνάρτηση κόστους. Ένας από αυτούς του αλγορίθμους είναι ο gradient descent. Η αρχικοποίηση των τιμών του W, b δεν θα πρέπει σε καμία περίπτωση να είναι ομοιόμορφη δίνοντας ως αρχική τιμή σε όλες τις παραμέτρους το 0. Αντίθετα, θα πρέπει να αρχικοποιηθούν με τυχαίες τιμές. Κάθε επανάληψη του gradient descent θα ανανεώνει τις παραμέτρους όπου α ο ρυθμός εκμάθησης.

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b)$$

Για την εύρεση αυτών των παραγώγων χρησιμοποιείται ο αλγόριθμος οπισθοδιάδοσης. Σαν πρώτο βήμα του αλγορίθμου εκτελείται αρχικά ένα πέρασμα προς τα εμπρός

μέχρι και τον υπολογισμό των ενεργοποιήσεων κάθε επιπέδου μέχρι και της συνάρτησης $h_{W,b}$. Για κάθε επίπεδο l υπολογίζεται ένας δείκτης σφάλματος $\delta_i^{(l)}$ ο οποίος ορίζει την συμμετοχή κάθε νευρώνα στο σφάλμα.

$$\delta_i^{(n_l)} = \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2} \left\| y - h_{W,b}(x) \right\|^2$$

Για $l = n_l - 1, n_l - 2, ...2$ για κάθε νευρώνα i στο επίπεδο l

Θέτουμε $\delta_i^{(l)} = \sum_{j=1}^{s_{l+1}} W_{ij}^{(l)} \delta_j^{(l+1)} f'(z_i^{(l)})$

Τέλος υπολογίζουμε τις μερικές παραγώγους ως εξής

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b : x, y) = \alpha_j^{(l)} \delta_i^{(l+1)}$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b : x, y) = \delta_i^{(l+1)}$$

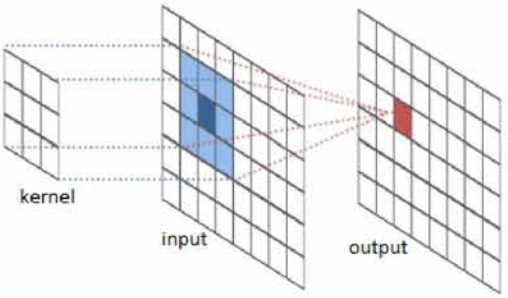
4.4 Νευρωνικά δίκτυα συνέλιξης (ΝΔΣ)

Τα ΝΔΣ έχουν αναπτυχθεί τις τελευταίες τρεις δεκαετίες και καθιερώθηκαν ως ένα ισχυρό εργαλείο κατηγοριοποίησης σε προβλήματα υπολογιστικής όρασης. Η επιτυχία τους πιστώνεται στην δυνατότητα που παρουσιάζουν να αντιλαμβάνονται αναπαραστάσεις αντικειμένων, προσώπων, χαρακτήρων σε σύγκριση με την σχεδίαση χειρωνακτικά κάποιων χαρακτηριστικών χαμηλού επιπέδου, τεχνική που χρησιμοποιείται από άλλους αλγορίθμους. Ωστόσο, η εκπαίδευση ενός τέτοιου δικτύου απαιτεί το δίκτυο να τροφοδοτηθεί με μεγάλο όγκο πληροφορίας και να υπολογιστούν πολλοί παράμετροι. Τα βασικά συστατικά και οι ιδιότητες ενός ΝΔΣ είναι: τοπικό πεδίο υποδοχής, η διαμοίραση των βαρών, η πράξη της συνέλιξης, τοπική συγκέντρωση, dropout, αλγόριθμοι βελτιστοποίησης.

4.4.1 Το επίπεδο της συνέλιξης

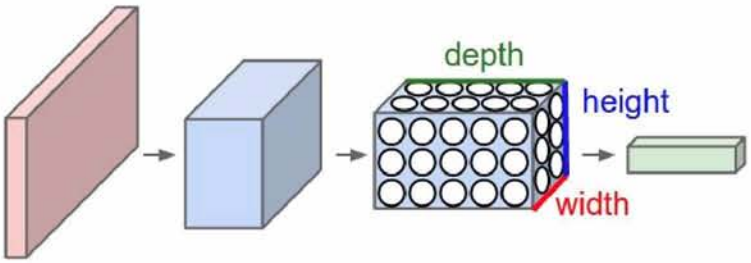
Τα ΝΔΣ λειτουργούν ως εξαγωγείς χαρακτηριστικών και με αυτό τον τρόπο μαθαίνουν τις αναπαραστάσεις μοντέλων που απεικονίζονται στις εικόνες. Έχουν την

δυνατότητα να μάθουν να αντιλαμβάνονται γωνίες, τελείες, μαύρα και λευκά σημεία και άλλα πολλά χαρακτηριστικά. Το επίπεδο της συνέλιξης αποτελεί το μέρος του δικτύου που καταναλώνεται η περισσότερη υπολογιστική ισχύς. Η πράξη της συνέλιξης συμβαίνει μεταξύ ενός διανύσματος μιας εικόνας, το οποίο στη συγκεκριμένη περίπτωση πρόκειται για ένα διάνυσμα δύο διαστάσεων καθώς οι εικόνες που χρησιμοποιούνται αποτελούνται μόνο από αποχρώσεις του γκρι και ενός φίλτρου μικρού μεγέθους συνήθως ($3\times3, 4\times4, 5\times5$). Στην περίπτωση που το διάνυσμα εισόδου είναι μεγαλύτερης διάστασης το φίλτρο θα πρέπει να έχει το ίδιο βάθος. Κάθε φίλτρο που χρησιμοποιείται είναι υπεύθυνο για να εξάγει ένα διαφορετικό χαρακτηριστικό από τα δεδομένα εισόδου. Κατά την πράξη της συνέλιξης το φίλτρο υπολογίζει ουσιαστικά το εσωτερικό γινόμενο μεταξύ της καταληφθείσας περιοχής ανάμεσα σε αυτό και στα δεδομένα εισόδου. Μετά από κάθε υπολογισμό μετατοπίζεται πάνω στην εικόνα με ένα σταθερό βήμα που έχει οριστεί και επαναλαμβάνεται αυτή η διαδικασία μέχρι να καλυφθεί το σύνολο της εικόνας.



Σχημα 4.2: Η πράξη της συνέλιξης (σχήμα από [34]).

Το τελικό αποτέλεσμα αυτής της διαδικασίας αποτελεί έναν χάρτη χαρακτηριστικών ή feature map. Κάθε φίλτρο παράγει και έναν διαφορετικό χάρτη ανεξάρτητο από τους υπόλοιπους. Αυτό επαναλαμβάνεται για έναν σχετικό μεγάλο αριθμό φίλτρων παράγοντας αντίστοιχα τους χάρτες χαρακτηριστικών και όλοι μαζί στοιβαγμένοι αποτελούν την έξοδο του επιπέδου. Τα πολλαπλά επίπεδα ΝΔΣ βοηθούν στην εκμάθηση πιο πολύπλοκα αναγνωρίσιμων μοντέλων και αναπαραστάσεων. Εμπειρικά, διαπιστώνεται πως τα ΝΔΣ έχουν μεγαλύτερη αποτελεσματικότητα και αναγνωρίζουν πιο πολύπλοκα μοτίβα, όταν κατά την μεταπήδηση σε βαθύτερα επίπεδα στο δίκτυο ο αριθμός των φίλτρων που χρησιμοποιούνται αυξάνεται, άρα και η μορφοποίηση των εξαγόμενων δεδομένων έχει μεγαλύτερο βάθος. Μία ενδεικτική απεικόνιση των δεδομένων σε κάθε επίπεδο ενός ΝΔΣ παρουσιάζεται παρακάτω.



Σχημα 4.3: Απεικόνιση της μορφοποίησης των δεδομένων μέσα σε ενα ΝΔΣ (σχήμα από [35]).

Η πράξη της συνέλιξης πραγματοποιείται μεταξύ του διανύσματος εισόδου και του φίλτρου. Όταν το μέγεθος της εικόνας έχει τα χαρακτηριστικά $W_1 \times H_1 \times D_1$ με παραμέτρους:

- το βάθος του διανύσματος εισόδου D
- το βήμα S
- το μέγεθος του φίλτρου F

4.4.1.1 Βήμα και γέμισμα μηδενικών

Δυο από τις παραμέτρους που ρυθμίζουν την λειτουργία ενός συνελικτικού επιπέδου είναι το βήμα ή stride και το γέμισμα μηδενικών ή zero-padding. Το βήμα αντιπροσωπεύει των αριθμό των pixels που μετακινείται το φίλτρο κάθετα και οριζόντια πάνω στην εικόνα κατά την διάρκεια της συνέλιξης. Το παραγόμενο κομμάτι δεδομένων μετά την πράξη της συνέλιξης παίρνει την μορφή $W_2 \times H_2 \times D_2$ με διαστάσεις

$$W_2 = (W_1 - F)/S + 1$$
$$H_2 = (H_1 - F)/S + 1$$
$$D_2 = D_1$$

Το μέγεθος του βήματος αποτελεί μια υπερ-παράμετρο η οποία όμως υπακούει σε κάποιους περιορισμούς. Δεν είναι εφικτές οι επιλογές όλων των τιμών βήματος. Αντίθετα η επιλογή του βήματος εξαρτάται από το μέγεθος της εικόνας. Για παράδειγμα σε μία εικόνα με μέγεθος 10×10 και φίλτρο με μέγεθος 3×3 είναι αδύνατον το βήμα να έχει την τιμή $S = 2$. Σε αυτή την περίπτωση κατά την πράξη της συνέλιξης το φίλτρο δεν ταιριάζει χωρικά πάνω στην εικόνα εισόδου για να ολοκληρωθεί η

πράξη. Για αυτό τον λόγο, χρησιμοποιείται μια επιπρόσθετη λειτουργία, το γέμισμα της εικόνας εισόδου.

Η παράμετρος padding ή zero-padding εφόσον το γέμισμα αφορά την προσθήκη μηδενικών, πρόκειται για την προσθήκη μηδενικών στα σύνορα της εικόνας εισαγωγής. Το μέγεθος αυτής της πράξης αποτελεί μια υπερ-παράμετρο. Με την χρήση αυτής της παραμέτρου μπορούμε να ρυθμίσουμε το μέγεθος του εξαγόμενου τόμου δεδομένων που προκύπτει από την πράξη της συνέλιξης. Χρησιμοποιείται κυρίως για να επιτρέψει το μέγεθος των εξαγόμενων δεδομένων να συμπίπτει σε ύψος και πλάτος με τα δεδομένα εισόδου.

4.4.1.2 Τοπική συνεκτικότητα

Μια από τις ιδιότητες που χαρακτηρίζει τα ΝΔΣ είναι η τοπική συνεκτικότητα. Κάθε μονάδα ενός επιπέδου του ΝΔΣ συνδέεται με μια υποπεριοχή της συνολικής εικόνας. Αυτή η υποπεριοχή αποτελεί την περιοχή εστίασης και έχει τις διαστάσεις και το βάθος του φίλτρου που χρησιμοποιείται. Αυτή η ιδιότητα εξυπηρετεί σε θέματα υπολογιστικής διαχείρισης και κόστους. Στην περίπτωση που κάθε μονάδα ενός επιπέδου συνδεόταν με όλες τις μονάδες του προηγούμενου επιπέδου θα ήταν υπολογιστικά ακριβή και μη διαχειρίσιμη η εκτέλεση του αλγορίθμου. Με αυτό τον τρόπο περιορίζεται ο αριθμός των παραμέτρων και μετριάζεται το φαινόμενο της υπερπροσαρμογής.

4.4.1.3 Κοινή χρήση παραμέτρων

Η κοινή χρήση των παραμέτρων πραγματοποιείται με στόχο τον περιορισμό του αριθμού των παραμέτρων στο δίκτυο. Ο κάθε χάρτης χαρακτηριστικών μοιράζεται τα ίδια βάρη καθώς είναι αποτέλεσμα της συνέλιξης της εικόνας με το ίδιο φίλτρο. Με αυτό τον τρόπο, ουσιαστικά μπορεί να εντοπιστεί το ίδιο χαρακτηριστικό σε όλες τις πιθανές θέσεις της εικόνας. Στον παρακάτω πίνακα παρουσιάζεται ο αριθμός των παραμέτρων όταν υπάρχει φίλτρο διαστάσεων $(10 \times 10 \times 3)$ και το αποτέλεσμα της συνέλιξης έχει διαστάσεις $(32 \times 32 \times 16)$.

Παρατηρούμε πως στην περίπτωση που δεν μοιράζονται οι παράμετροι και κάθε στοιχείο της εικόνας χρησιμοποιεί έναν διαφορετικό πίνακα βαρών, ο αριθμός των παραμέτρων εκτοξεύεται σε έναν μη διαχειρίσιμο αριθμό.

	Με κοινή χρήση	**Χωρίς κοινή χρήση**
Νευρώνες	$32 \times 32 = 1024$	$32 \times 32 \times 16 = 16384$
Σύνολο Παραμέτρων	$1024 \times 332 = 339968$	$16384 \times 332 = 5439488$

ΠΙΝΑΚΑΣ 4.1: Υπολογισμός παραμέτρων στην περίπτωση που μοιράζονται παράμετροι και σε αυτήν που δεν μοιράζονται.

4.4.2 Συγκέντρωση

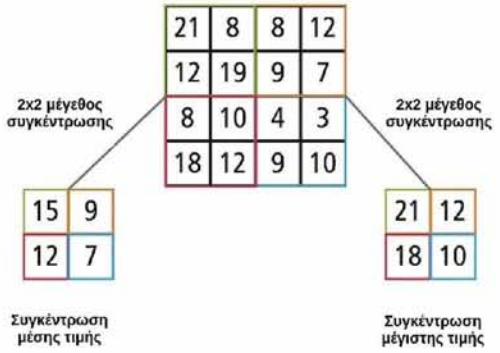
Η συγκέντρωση [36] ή pooling πρόκειται για μια τεχνική δειγματοληψίας που στοχεύει στο να να περιορίσει το μέγεθος μιας αναπαράστασης και να την καταστήσει πιο εύκολα διαχειρίσιμη. Το δίκτυο δέχεται ως είσοδο έναν συγκεκριμένο όγκο δεδομένων και τον συμπιέζει χωρικά. Ουσιαστικά, ένα κομμάτι χωρικής πληροφορίας πετιέται με σκοπό τον εντοπισμό κάποιων ιδιαίτερων χαρακτηριστικών (μια γωνία, μια καμπύλη) αγνοώντας την θέση του αντικειμένου στην εικόνα. Η συγκεκριμένη πράξη πραγματοποιείται σε κάθε περιοχή ενεργοποίησης ξεχωριστά. Αυτό είναι σημαντικό για την αναγνώριση όμοιων αντικειμένων που μπορεί να βρίσκονται σε διαφορετικές θέσεις σε κάθε εικόνα.

Υπάρχουν αρκετά είδη συγκέντρωσης. Δύο από αυτά είναι το μέγιστο και αυτό της μέσης τιμής. Στην μέγιστη συγκέντρωση επιλέγεται την μέγιστη τιμή της επιλεγμένης περιοχής ενώ η συγκέντρωση μέσης τιμής επιλέγει την μέση τιμή της επιλεγμένης περιοχής.

Η τεχνική της συγκέντρωσης μπορεί να οδηγήσει σε μειωμένη απόδοση του αλγορίθμου εάν το μέγεθος του πίνακα συγκέντρωσης είναι πολύ μεγάλο.

Αντίστοιχα με την πράξη της συνέλιξης, στην πράξη της συγκέντρωσης, το επίπεδο δέχεται ένα κομμάτι δεδομένων με μέγεθος $W_1 \times H_1 \times D_1$ και το συμπιέζει σε $W_2 \times H_2 \times D_2$ με τον ίδιο τρόπο. Για παράδειγμα, με την χρήση πίνακα συγκέντρωσης (2×2) που χρησιμοποιείται στην εργασία παρακάτω και το βήμα να λαμβάνει την τιμή 2, το μέγεθος των δεδομένων περιορίζεται στο μισό.

Στην εικόνα 4.4 υπάρχει παράδειγμα μέγιστης συγκέντρωσης και συγκέντρωσης μέσης τιμής (4×4) με μέγεθος πίνακα συγκέντρωσης (2×2) και βήμα 2.



ΣΧΗΜΑ 4.4: Συγκέντρωση μέγιστης και μέσης τιμής.

4.4.3 Τεχνικές αποφυγής υπερπροσαρμογής

4.4.3.1 Κανονικοποίηση

Υπάρχουν αρκετοί τρόποι για τον έλεγχο της χωρητικότητας ενός ΝΔ για την αποφυγή της υπερπροσαρμογής. Η L2 κανονικοποίηση ή L2 regularization πρόκειται ίσως για την πιο διαδεδομένη τεχνική. Η υλοποίηση της βασίζεται στον περιορισμό της δυναμικής των βαρών που λαμβάνουν πολύ υψηλές τιμές. Αντίθετα, θεωρείται προτιμότερο να συμβάλλουν όλες οι παράμετροι σε μικρότερο βαθμό ορίζοντας έναν συντελεστή $\frac{1}{2}\lambda w^2$, με τιμωρητικό ουσιαστικά χαρακτήρα προς τις μεγάλες τιμές βαρών, όπου το λ σηματοδοτεί την δυναμική της κανονικοποίησης. Ο συντελεστής $1/2$ εξυπηρετεί έτσι ώστε το τελικό μέγεθος της κλήσης να έχει την τιμή λw, αντί $2\lambda w$.

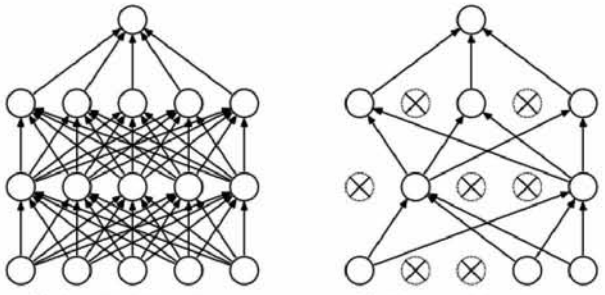
Πέρα από την τεχνική της L2 κανονικοποίησης υπάρχει και η L1 κανονικοποίηση. Πρόκειται, για μια παραλλαγή [37] κατά την οποία ο συντελεστής τώρα λαμβάνει την τιμή $\lambda|w|$ σε κάθε τιμή των βαρών. Με αυτό τον τρόπο επιτυγχάνεται να γίνει χρήση ενός υποσυνόλου των εισόδων καθιστώντας το δίκτυο ανεκτικό σε 'θορυβώδεις' εισόδους.

Υπάρχει η δυνατότητα εφαρμογής αυτών των 2 τεχνικών της L1 κανονικοποίησης και L2 κανονικοποίησης ταυτόχρονα με τον συντελεστή να παίρνει την μορφή $\lambda_1|w| + \lambda_2 w^2$.

4.4.3.2 Dropout

Το dropout [38] πρόκειται για μια τεχνική που αντιμετωπίζει το φαινόμενο της υπερπροσαρμογής ή αλλιως overfitting. Το φαινόμενο αυτό συμβαίνει όταν ο αλγόριθμος δεν γενικεύει καλά και δεν είναι αποδοτικός όταν χρησιμοποιούνται δεδομένα που δεν έχει "δει". Αυτό σημαίνει πως παρουσιάζεται μια έλλειψη ακρίβειας στα αποτελέσματα που προέρχονται από δεδομένα δοκιμής σε σχέση με τα δεδομένα εκπαίδευσης. Η ιδέα πίσω από αυτή την μεθοδολογία είναι η απενεργοποίηση κάποιων τυχαίων μονάδων του νευρωνικού δικτύου. Αυτό το γεγονός αποτρέπει την συμπροσαρμογή των δεδομένων αναγκάζοντας έτσι κάθε νευρώνα να εντοπίζει χαρακτηριστικά από προηγούμενα επίπεδα, ανεξάρτητα από τους νευρώνες τους ίδιου επιπέδου. Αποτελεί ένα είδος κανονικοποίησης του δικτύου προσθέτοντας θόρυβο στο δίκτυο με αποτέλεσμα ο αλγόριθμος να γενικεύει καλύτερα ειδικά σε προβλήματα όρασης. Κατά την εκτέλεση του dropout κάποιες μονάδες απενεργοποιούνται με πιθανότητα p που ισούται συνήθως με 0.5.

Στην εικόνα 4.5 εμφανίζεται ένα παράδειγμα εφαρμογής της τεχνικής dropout σε ένα ΝΔ με 2 κρυφά επίπεδα.



Σχημα 4.5: Παράδειγμα dropout (σχήμα από [38]).

4.4.4 Συναρτήσεις ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης χρησιμοποιούνται για να καθορίσουν αν θα ενεργοποιηθεί κάποια μονάδα του ΝΔ και αν θα περάσει σήμα μέσα από αυτό. Οι συναρτήσεις που χρησιμοποιούνται κυρίως στα ΝΔ είναι μη γραμμικές και μετατρέπουν το σήμα

που περνάει από αυτές από γραμμικό σε μη γραμμικό. Οι πιο συνηθισμένες συναρτήσεις ενεργοποίησης είναι: η σιγμοειδής, η υπερβολική εφαπτομένη, η ReLU και η leaky ReLU.

4.4.4.1 Ανορθωμένη γραμμική μονάδα (ReLU)

Η συνάρτηση ReLU έχει αποκτήσει ιδιαίτερη σημασία την τελευταία δεκαετία, ιδιαίτερά στα ΝΔΣ. Είναι μία απλή μη γραμμική συνάρτηση ενεργοποίησης που ενεργοποιεί μόνο τις θετικές τιμές. Θεωρείται ελκυστική επειδή επιταχύνει τον στοχαστικό αλγόριθμο απότομης καθόδου συγκριτικά με την σιγμοειδή και την εφαπτομένη. Δεν είναι υπολογιστικά ακριβή και έχει την μορφή

$$f(x) = max(0, x)$$

4.4.4.2 Σιγμοειδής

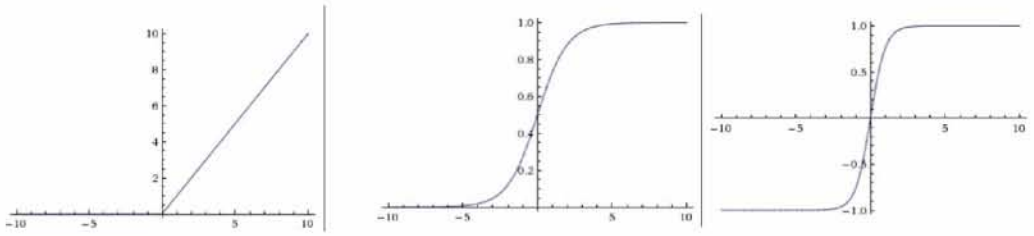
Συμπτύσσει τις τιμές που δέχεται μεταξύ του [0,1]. Η σιγμοειδής συνάρτηση (sigmoid) δεν προτιμάται τόσο συχνά τελευταία επειδή διακρίνεται από κάποια σημαντικά ελαττώματα. Αρχικά, έρχεται σε κορεσμό εύκολα, εξαφανίζει την κλίση και το δίκτυο δεν μπορεί να εκπαιδευτεί αποδοτικά. Το άλλο μειονέκτημα της είναι πως δεν πρόκειται για μια συνάρτηση που έχει ως κέντρο το σημείο 0. Έτσι στην περίπτωση που τα δεδομένα είναι όλα αρνητικά ή θετικά, ο αλγόριθμος της οπισθοδιάδοσης δεν μπορεί να συγκλίνει. Τέλος, θεωρείται υπολογιστικά ακριβή συνάρτηση. Η σιγμοειδής συνάρτηση έχει την μορφή:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

4.4.4.3 Υπερβολική εφαπτομένη

Ο υπερβολική εφαπτομένη (tanh) περιορίζει τις τιμές που την διαπερνούν μεταξύ [-1,1]. Είναι παρόμοια συνάρτηση με την σιγμοειδή, έρχεται και αυτή σε κορεσμό εξαφανίζοντας την κλίση με μοναδική διαφορά πως η συγκεκριμένη έχει ως κέντρο το 0 και για αυτό προτιμάται. Η μαθηματική της μορφή είναι:

$$\tanh(x) = 2\sigma(2x) - 1$$



ΣΧΗΜΑ 4.6: Συναρτήσεις ReLU, σιγμοειδή και υπερβολική εφαπτομένη.

4.4.5 Συνάρτηση κόστους

Η συνάρτηση απωλειών υπολογίζει την απόκλιση που εμφανίζεται μεταξύ των προβλέψεων του αλγορίθμου μας και του πίνακα αλήθειας. Αυτή η συνάρτηση ορίζεται ως ο μέσος όρος απωλειών από κάθε παράδειγμα ξεχωριστά $Loss = \frac{1}{N}\sum_i^N L_i$, όπου N ορίζεται το σύνολο των δειγμάτων και ως L_i η απώλειες κάθε δείγματος. Στην περίπτωση που η συνάρτηση ενεργοποίησης του επιπέδου εξόδου είναι της μορφής $f(x_i : W)$. Σε προβλήματα κατηγοριοποίησης, όπου είναι της μορφής ένα εναντίον όλων, μία από τις πιο διαδεδομένες συναρτήσεις μέτρησης του κόστους L_i είναι το hinge loss που χρησιμοποιείται στον αλγόριθμο ΜΔΣ.

$$L_i = \sum_{j \neq y_i} max(0, f_j - f_{y_i} + 1)$$

Η εναλλακτική επιλογή συνάρτησης κόστους είναι ο softmax ταξινομητής που χρησιμοποιεί ως μέτρηση απωλειών το cross-entropy όπως παρουσιάζεται παρακάτω.

Ο softmax ταξινομητής πρόκειται για μια γενίκευση του αλγορίθμου logistic regression για την κατηγοριοποίηση σε δυο κλάσεις που αντιστοιχεί σε πολλές κλάσεις. Σε αντίθεση με το hinge loss που παρουσιάζεται κατά την κατηγοριοποίηση μέσω του αλγορίθμου ΜΔΣ, το cross-entropy είναι η αντίστοιχη συνάρτηση κόστους στον softmax ταξινομητή. Η συνάρτηση softmax λαμβάνει πραγματικές τιμές και τις περιορίζει στο διάστημα [0,1] όπου τις εκφράζει σε μορφή πιθανοτήτων. Η μαθηματική μορφή είναι:

$$f_j(z) = \frac{e^{f_{y_j}}}{\sum_k e^{f_k}}$$

Το cross-entropy μεταξύ μιας πραγματικής κατανομής p και μιας υπολογισμένης κατανομής q ορίζεται ως:

$$H(p,q) = -\sum_x p(x) \log q(x)$$

Ο softmax ταξινομητής τείνει να ελαχιστοποιήσει το cross-entropy μεταξύ της υπολογισμένης πιθανότητας της κλάσης $q = \frac{e^{f_{y_i}}}{\sum_j e^{f_j}}$ και της πραγματικής πιθανότητας, η οποία ισοδυναμεί με ένα διάνυσμα [0..1...0] με μοναδικό 1 στην θέση y_i που είναι η θέση της σωστής κλάσης (one hot vector).

4.4.6 Αλγόριθμοι βελτιστοποίησης

Οι αλγόριθμοι βελτιστοποίησης χρησιμοποιούνται με στόχο να ελαχιστοποιήσουν ή να μεγιστοποιήσουν την επικείμενη συνάρτηση (συνάρτηση σφάλματος, E(x)). Η συνάρτηση σφάλματος πρόκειται για μία μαθηματική συνάρτηση η οποία έχει ως ορίσματα τις εσωτερικές παραμέτρους του μοντέλου του ΝΔ. Αυτές οι παράμετροι, επιδέχονται αλλαγή, εκπαιδεύονται και προσαρμόζονται στις απαιτήσεις του προβλήματος με στόχο την εύρεση της βέλτιστης λύσης. Τέτοιες παράμετροι είναι για παράδειγμα τα βάρη (weights) και η τιμή bias .

Η ελαχιστοποίηση της συνάρτησης σφάλματος παίζει καθοριστικό ρόλο κατά την διαδικασία της εκπαίδευσης του ΝΔ. Οι εσωτερικές παράμετροι συμβάλλουν καθοριστικά στην ακρίβεια και την αποδοτικότητα του ΝΔ. Για αυτό τον λόγο έχουν αναπτυχθεί διάφοροι αλγόριθμοι που ρυθμίζουν και ανανεώνουν τις τιμές των παραμέτρων με στόχο την εύρεση των βέλτιστων τιμών. Κάποιοι από αυτούς του αλγόριθμους που χρησιμοποιούνται στα ΝΔ είναι ο αλγόριθμος σύγκλισης με ελάττωσης της παραγώγου, Adagrad, Adadelta [39], Adam [40], RMSprop [41].

4.4.6.1 Αλγόριθμος σύγκλισης με ελάττωση της παραγώγου

Ο αλγόριθμος σύγκλισης με ελάττωση της παραγώγου ή gradient descent πρόκειται για τον πιο γνωστό αλγόριθμο βελτιστοποίησης που αποτελεί την βάση για το πως

εκπαιδεύουμε και ρυθμίζουμε τα ΝΔ. Η μαθηματική έκφραση την οποία ακολουθεί ο αλγόριθμος για την ανανέωση των παραμέτρων είναι μέχρι να επιτευχθεί η σύγκλιση είναι:

$$\theta = \theta - \alpha * \nabla J(\theta)$$

Όπου το α είναι ο ρυθμός εκμάθησης και η συνάρτηση $J(\theta)$ είναι η συνάρτηση του σφάλματος που επιδιώκουμε να ελαχιστοποιήσουμε. Η παράμετρος που ρυθμίζεται σε αυτό τον αλγόριθμο είναι κατά κύριο λόγο τα βάρη. Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται κατά την εκτέλεση του αλγορίθμου της οπισθοδιάδοσης με κατεύθυνση προς τα πίσω, χρησιμοποιεί τις πληροφορίες που διαθέτει για το σφάλμα της συνάρτησης έτσι ώστε να ανανεώσει τις τιμές των βαρών για να ελαχιστοποιηθεί το λάθος.

Ο αλγόριθμος gradient descent εμφανίζεται και σε άλλες μορφές εκτός από την απλή του. Υπάρχει ο στοχαστικός αλγόριθμος σύγκλισης με ελάττωση της παραγώγου ή SGD, mini-batch gradient descent, nesterov accelerated gradient.

4.4.6.2 Adagrad

Ο αλγόριθμος Adagrad απλά επιτρέπει να προσαρμόζεται ο ρυθμός εκμάθησης α πάνω στις παραμέτρους. Χρησιμοποιεί διαφορετικό ρυθμό εκμάθησης για κάθε παράμετρο βασισμένος στην παράγωγο που υπολογίστηκε παλιότερα για την συγκεκριμένη παράμετρο. Η μαθηματική φόρμα αυτού του αλγορίθμου είναι :

$$\theta_{t+1,i} = \theta_{t,i} - \frac{a}{\sqrt{G_{t,ii} + \epsilon}} g_{t,i}$$

Ο αλγόριθμος adagrad ρυθμίζει τον ρυθμό εκμάθησης α σε κάθε βήμα t για κάθε παράμετρο $\theta(i)$ βασισμένος σε παλιότερες τιμές της παραγώγου. Το g αποτελεί την κλίση της συνάρτησης σφάλματος και το G είναι ένας διαγώνιος πίνακας που στην διαγώνιο περιλαμβάνει τις τιμές του τετραγώνου της παραγώγου.

Ένα από τα πλεονεκτήματα αυτού του αλγορίθμου είναι πως δεν χρειάζεται να ρυθμιστεί χειροκίνητα ο ρυθμός εκμάθησης. Οι περισσότερες υλοποιήσεις δίνουν μια σταθερή τιμή 0.01. Το αρνητικό είναι πως ο ρυθμός εκμάθησης μειώνεται πάντα

καθώς όλες οι τιμές που προστίθενται είναι θετικές. Έτσι μειώνεται συνεχώς η ταχύτητα εκπαίδευσης του ΝΔ.

4.4.6.3 Adadelta

Πρόκειται για μία επέκταση του αλγορίθμου adagrad ξεπερνώντας το πρόβλημα της συνεχόμενης μείωσης του ρυθμού εκμάθησης. Αντίθετα, περιορίζει αυτή την μείωση των συγκεντρωμένων τιμών της παραγώγου σε μια σταθερή τιμή. Αντί να υπολογίζει το άθροισμα των τετραγώνων των τιμών της παραγώγου κρατάει την μέση τιμή τους. Η τρέχουσα μέση τιμή είναι:

$$E[g^2]_t = \gamma E[g]_{t-1} + (1-\gamma){g_t}^2$$

Θέτουμε το γ μια τιμή περίπου 0.9. Για καλύτερη σαφήνεια γράφουμε την ανανέωση του SGD σε όρους της παραμέτρους $\Delta\theta_t$

$$\Delta\theta_t = -\alpha g_{t,i}$$

$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

Η παράμετρος ανανέωσης από τον αλγόριθμο adadelta παίρνει επομένως την μορφή:

$$\Delta\theta_t = -\frac{a}{\sqrt{G_t + \epsilon}} \odot g_t$$

Αντικαθιστούμε τον διαγώνιο πίνακα G με την μέση τιμή των τετραγώνων της παραγώγου $E[g^2]$

$$\Delta\theta_t = -\frac{a}{\sqrt{E[g^2]_t + \epsilon}} \odot g_t$$

Εφόσον ο παρονομαστής είναι απλά η τιμή του RMS μπορούμε να αντικαταστήσουμε την τιμή

$$\Delta\theta_t = -\frac{a}{RMS[g]_t} g_t$$

4.4.6.4 RMSprop

Ο αλγόριθμος RMSprop πρόκειται για μια προσπάθεια που έγινε από τον Geoffrey Hinton και αναπτύχθηκε σε παράλληλη χρονική περίοδο με τον αλγόριθμο Adagrad για να αντιμετωπιστεί το πρόβλημα που αντιμετώπιζε ο αλγόριθμος Adadelta με την συνεχόμενη πτώση του ρυθμού εκμάθησης. Είναι ουσιαστικά πανομοιότυπη προσπάθεια με την πρώτη ανανέωση του διανύσματος του αλγορίθμου Adadelta που περιγράψαμε παραπάνω. Ο Hinton προτείνει ως τιμή της παραμέτρου $\gamma = 0.9$ και του ρυθμού εκμάθησης $\alpha = 0.001$

$$E[g^2]_t = 0.9E[g]_{t-1} + 0.1{g_t}^2$$

$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

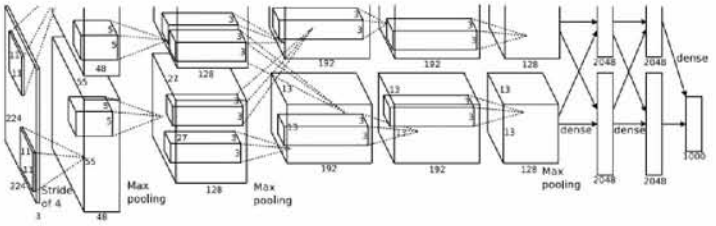
4.4.7 Διάσημα μοντέλα ΝΔΣ

Στο παρελθόν έχουν δημιουργηθεί αρκετές αρχιτεκτονικές ΝΔΣ που αποτέλεσαν ορόσημο και έθεσαν τις βάσεις για περαιτέρω ανάπτυξη στον τομέα της υπολογιστικής όρασης:

LeNet: Αναπτύχθηκε το 1990 από τον Yann LeCun και πρόκειται για την πρώτη επιτυχημένη εφαρμογή των ΝΔΣ. Χρησιμοποιήθηκε στην αναγνώριση κωδικών και ψηφίων. Ουσιαστικά, αυτή η έρευνα σηματοδότησε το έναυσμα για την ένταξη της αυτόματης μάθησης στον τομέα των συστημάτων αναγνώρισης προτύπων.

AlexNet [42]: αναπτύχθηκε από τους Alex KrizhevskyIlya, Sutskever, Geoff Hinton και συμμετείχε στον διαγωνισμό ImageNet ILSVRC challenge το 2012 ξεπερνώντας τον δεύτερο φιναλίστ με μεγάλη διαφορά. Αυτό το μοντέλο είχε παρόμοια αρχιτεκτονική με το LeNet, με μόνες διαφορές πως αποτελούνταν από περισσότερα επίπεδα και εισήγαγε ουσιαστικά την πολυεπίπεδη συνέλιξη.

GoogLeNet: αναπτύχθηκε από τον Szegedy και άλλους [43], από την Google το 2014 και αναδείχθηκε νικητής του διαγωνισμού ILSVRC 2014. Η κύρια συνεισφορά του στον τομέα ήταν το Inception Module που βοήθησε να μειωθεί ο αριθμός των



Σχημα 4.7: Αναπαράσταση της AlexNet αρχιτεκτονικής (σχήμα από [42]).

παραμέτρων στο δίκτυο συγκριτικά με το AlexNet από 60M στα 4M, καθώς επίσης και στην χρησιμοποίηση του επιπέδου συγκέντρωσης μέσης τιμής αντί από ένα πλήρως συνδεδεμένο επίπεδο στην κορυφή του ΝΔΣ.

Την δεύτερη θέση σε αυτό τον διαγωνισμό κέρδισε το μοντέλο **VGGNet** απο τους Karen Simonyan, Andrew Zisserman [44]. Μέσα από την έρευνα τους έδειξαν ότι το βάθος του δικτύου συνδέεται στενά με την αποδοτικότητα. Το μοντέλο τους περιλαμβάνει 16 συνελικτικά και πληρως συνδεδεμένα επίπεδα και μία αρχιτεκτονική που περιλαμβάνει μόνο 3×3 συνελικτικά φίλτρα και 2×2 επίπεδα συγκέντρωσης.

Το **ResNet** αναπτύχθηκε από τον Kaiming He και άλλους, και κέρδισε τον διαγωνισμό ILSVRC 2015.

Κεφάλαιο 5

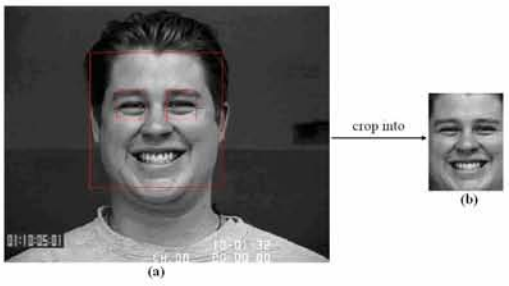
Προγραμματιστικό Μέρος

Το προγραμματιστικό μέρος της εργασίας για την δημιουργία ενός ΝΔ ικανού να αναγνωρίζει συναισθήματα μέσα από βίντεο βασίστηκε στις βάσεις δεδομένων CK+, FER2013 για την εκπαίδευση του ΝΔ, ενώ ο έλεγχος τον αποτελεσμάτων πραγματοποιήθηκε στην ΒΔ EMOTIC [22]. Παίρνοντας ως δεδομένο πως ένα βίντεο αποτελείται ένα σύνολο συνεχόμενων εικόνων (πλάνων), η αποστολή του ΝΔ που δημιουργήθηκε είναι να αναγνωρίζει το συναίσθημα από ανεξάρτητες εικόνες προσώπων. Η συνάρτηση λάθους που χρησιμοποιείται είναι η categorical cross-entropy. Παρακάτω παρουσιάζονται 3 αρχιτεκτονικές ΝΔΣ που υλοποιήθηκαν με την βοήθεια της χρήσης GPU για βέλτιστη ταχύτητα εκπαίδευσης. Τα στοιχεία από τα βοηθητικά εργαλεία, βιβλιοθήκες και το υλικό του υπολογιστή πάνω στο οποίο δημιουργήθηκε η εργασία βρίσκονται στον παρακάτω πίνακα:

API	Keras	Tensorflow	OpenCV
GPU	Nvidia GTX 770, 2GB Μνήμη		
Γλώσσα Προγραμματισμού	Python		

Πινακας 5.1: Βασικά στοιχεία για την υλοποίηση του προγράμματος.

5.1 Τεχνικές οπτικής προεπεξεργασίας



Σχημα 5.1: Εντοπισμός του προσώπου, αποκοπή αυτής της περιοχής σε εικόνες greyscale με διαστάσεις 48×48.

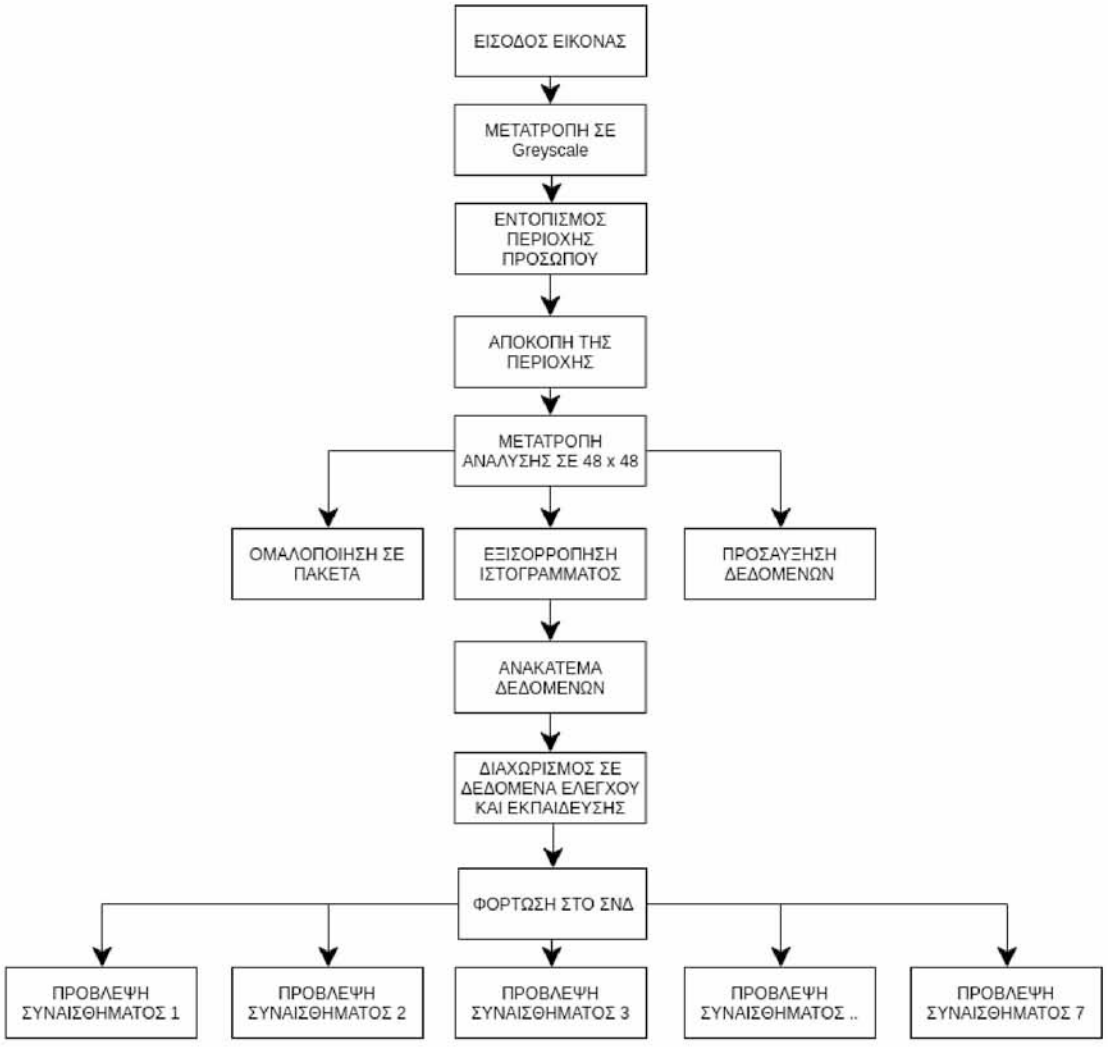
Η πρώτη φάση αυτής της εργασίας αποτελείται από τον εντοπισμό του προσώπου. Για αυτή την αποστολή χρησιμοποιήθηκε ο σειριακός ταξινομητής τύπου Haar, ο οποίος είναι υπεύθυνος για τον εντοπισμό της περιοχής που βρίσκονται τα χαρακτηριστικά του προσώπου σε μία εικόνα. Συγκεκριμένα, χρησιμοποιήθηκε ο Haar cascade frontal face default ταξινομητής, ο οποίος αποτελεί ένα προεκπαιδευμένο μοντέλο που παρέχεται από την βιβλιοθήκη OpenCV. Με αυτό τον τρόπο επιχειρούμε να μειώσουμε τον όγκο των δεδομένων αποκόπτοντας περιττή πληροφορία, η οποία ενδέχεται να προκαλέσει σύγχυση στο ΝΔΣ όταν θα την δεχτεί σαν είσοδο. Έτσι αυτή η περιοχή αφού εντοπιστεί γίνεται προσπάθεια να απομονωθεί από την υπόλοιπη εικόνα. Στη συνέχεια, επειδή το σύνολο των εικόνων που περιλαμβάνει η βάση δεδομένων CK+ είναι πολύχρωμες, επιχειρείται να μετατραπούν σε greyscale, ουσιαστικά μειώνοντας τις διαστάσεις από 3 (RGB) σε 1. Έπειτα, μετατρέπουμε το μέγεθος το εικόνων στην διάσταση 48×48 ύψος και πλάτος. Σε αυτό το σημείο ολοκληρώνεται η πρώτη φάση επεξεργασίας των δεδομένων, στην προσπάθεια που γίνεται να προετοιμαστεί η πληροφορία για να εισαχθεί στο ΝΔ, αφαιρώντας όσο το δυνατόν περισσότερη περιττή πληροφορία με απώτερο σκοπό να μειωθούν οι υπολογιστικές απαιτήσεις του αλγορίθμου που θα ακολουθήσουν και να επιτευχθεί η μέγιστη δυνατή ακρίβεια των αποτελεσμάτων. Στην παραπάνω εικόνα, απεικονίζεται όλη η διαδικασία αυτής της φάσης.

Το επόμενο βήμα για την προεπεξεργασία των εικόνων και την προετοιμασία τους για την είσοδο στο ΝΔΣ αποτελεί η εφαρμογή της τεχνικής εξισορρόπησης ιστογράμματος. Η τεχνική αυτή αυξάνει την αντίθεση που εμφανίζεται στις εικόνες όπως αναφέρεται στην παραπάνω ενότητα 3.1.1. Παρατηρήθηκε στο πρόγραμμά μας πως η εφαρμογή της τεχνικής της εξισορρόπησης ιστογράμματος μετά από την αλλαγή του μεγέθους της εικόνας είχε θετική επίδραση στην ακρίβεια του ΝΔ σε σύγκριση με την αντίστροφη σειρά εκτέλεσης.

Αξίζει να αναφερθεί πως όλες οι εικόνες είναι αποθηκευμένες σέ έναν πίνακα. Όλες οι εικόνες πριν εισαχθούν στο ΝΔΣ ανακατεύονται οι θέσεις που έχουν σε αυτόν

τον πίνακα για να επιτευχθεί καλύτερη κανονικοποίηση.

Μετέπειτα, χωρίζονται σε αναλογία 8 προς 2 οι εικόνες που θα χρησιμοποιηθούν για την εκπαίδευση προς αυτές που θα χρησιμοποιηθούν για έλεγχο.



Σχημα 5.2: Η σειρά των βημάτων που ακολουθήθηκαν στην εργασία.

5.2 Τοπολογία 1ου μοντέλου ΝΔΣ και αποτελέσματα

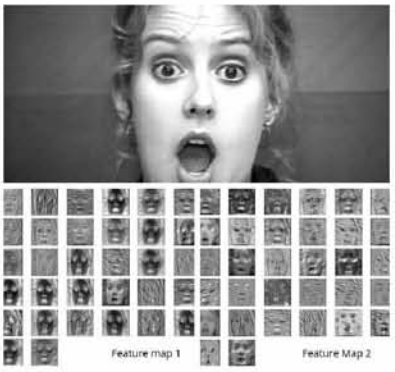
Σε αυτή την ενότητα περιγράφεται η τοπολογία των μοντέλων ΝΔΣ που εφαρμόστηκαν κατά την εκπαίδευση του ΝΔ και τα αποτελέσματα που επιτεύχθηκαν. Οι αξιολογήσεις των μοντέλων βασίστηκαν στις ΒΔ CK+ και FER2013. Η ΒΔ αποτελείται από 5,526 εικόνες προερχόμενες από την ΒΔ CK+ και 35,887 προερχόμενες από την ΒΔ FER2013. Από το σύνολο των εικόνων οι 33,130 χρησιμοποιούνται για την εκπαίδευση του ΝΔ και οι υπόλοιπες 8,282 για τον έλεγχο των αποτελεσμάτων. Όλες οι εικόνες που προέρχονται από την ΒΔ CK+ έχουν ληφθεί σε ελεγχόμενα εργαστήρια και τα υποκείμενα έχουν ποζάρει για κάποιο συγκεκριμένο συναίσθημα, ενώ οι εικόνες που προέρχονται απο την ΒΔ FER2013 προέρχονται από εικόνες του διαδικτύου. Με αυτό τον τρόπο επιτεύχθηκε η συλλογή δεδομένων με μεγάλο εύρος πληροφορίας. Παρ' όλο που η ΒΔ CK+ διαθέτει αρκετές εικόνες, τα δεδομένα επικεντρώνονται μόνο σε εικόνες και πλάνα με συγκεκριμένο ύφος, υπάρχει ελεγχόμενος φωτισμός, πρόσωπα ευθυγραμμισμένα στο κέντρο των εικόνων και γενικά πολύ ιδανικές συνθήκες που δεν επιτρέπουν το δίκτυο να εκπαιδευτεί και να ανταποκρίνεται σε πλάνα διαφορετικού ύφους. Για αυτό τον λόγο ενσωματώθηκε στην αρχική βάση CK+ μια ακόμα βάση δεδομένων η FER2013, η οποία διαθέτει μια ευρεία γκάμα εικόνων. Ο πρώτος στόχος ήταν να προσαρμοστεί η συγκεκριμένη βάση στα πλαίσια του αρχικού μας πλάνου, δηλαδή στον διαχωρισμό των εικόνων στις 7 βασικές κατηγορίες συναισθημάτων, καθώς η αρχική έκδοση της συγκεκριμένης ΒΔ συμπεριλαμβάνει μια επιπλέον κατηγορία, αυτήν της ουδέτερης έκφρασης. Επιπλέον το μέγεθος των εικόνων που δέχεται πλέον το ΝΔΣ είναι (48×48), καθώς αυτό είναι το μέγεθος των εικόνων που υποστηρίζει η βάση δεδομένων FER2013.

Το αποτέλεσμα που αποτελεί baseline για αυτή την ΒΔ προέρχεται από την χρήση του αλγορίθμου ΜΔΣ και το σκορ ακρίβειας που επιτευχθεί ανέρχεται στο 51.25%.

Ανάμεσα σε πολλές συναρτήσεις βελτιστοποίησης που δοκιμάστηκαν προτιμήθηκε ο adadelta με ρυθμό εκμάθησης $= 1$ $epsilon = None$. Στα πλαίσια αυτής της εργασίας, υλοποιήθηκε ένα μοντέλο ΝΔΣ με 3 συνελικτικά επίπεδα. Το πρώτο συνελικτικό επίπεδο δέχεται ως είσοδο έναν τόμο με διαστάσεις $(48 \times 48 \times 1)$ και πραγματοποιείται συνέλιξη με 32 φίλτρα, (3×3) διαστάσεων με βήμα 1. Στη συνέχεια ακολουθεί συνάρτηση ενεργοποίησης ReLU, χωρίς την ύπαρξη συνάρτησης συγκέντρωσης. Στο δεύτερο επίπεδο συνέλιξης χρησιμοποιούνται 32, (3×3) φίλτρα και έπειτα ακολουθεί μια συνάρτηση ενεργοποίησης ReLU, ένα επίπεδο συγκέντρωσης με (2×2)

μέγεθος και βήμα 2 και ένα επίπεδο dropout με τιμή 0.5. Στο τρίτο επίπεδο επίπεδο διπλασιάζεται ο αριθμός των φίλτρων στα 64 και προστίθεται ένα επίπεδο batch normalization.

Αυτό το επίπεδο βοήθησε το δίκτυο να ενισχύσει την σταθερότητα του και να επιταχύνει την διαδικασία εκπαίδευσης και να μειώσει τις απώλειες. Στη συνέχεια ακολουθεί ένα πλήρως συνδεδεμένο δίκτυο και ακολουθεί η συνάρτηση softmax για τον υπολογισμό των απωλειών. Οι υπερπαράμετροι που επιλέχθηκαν στο μοντέλο μας είναι αποτέλεσμα πολλών πειραματικών δοκιμών, με αρκετές παραλλαγές και προσπάθειες. Στις συγκεκριμένες τιμές το δίκτυο αντέδρασε με τον βέλτιστο τρόπο και απέδωσε την μεγαλύτερη ακρίβεια. Παρακάτω παρουσιάζονται λεπτομέρειες για την αρχιτεκτονική του μοντέλου ΝΔΣ 5.2. Παρακάτω παρουσιάζονται οι χάρτες χαρακτηριστικών για κάθε συνελικτικό επίπεδο, καθώς και η τοπολογία του μοντέλου μέσα από έναν πίνακα. Η συγκεκριμένη αρχιτεκτονική περιλαμβάνει συνολικά 438,439 παραμέτρους από τις οποίες 438,311 είναι εκπαιδεύσιμες και 128 σταθερές.



Σχημα 5.3: Εικόνα από την $CK+$ ΒΔ που εκφράζει την έκπληξη και απεικονίζει χάρτες χαρακτηριστικών.

Για την εκπαίδευση του μοντέλου χρησιμοποιήσαμε ως αλγόριθμο βελτιστοποίησης τον adadelta καθώς παρουσίασε τις καλύτερες επιδόσεις. Επιπλέον το μοντέλο εκπαιδεύτηκε σε batches = 256, epochs = 50.

Στην προσπάθεια να κατανοηθεί η λειτουργία του ΝΔΣ και τον τρόπο που αναγνωρίζει μοτίβα, εισάγαμε στο δίκτυο μία εικόνα 5.3 για ελέγξαμε τον τρόπο που αντιδρά το δίκτυο σε κάθε συνελικτικό επίπεδο. Η πρώτη διαπίστωση ήταν πως ο εντοπισμός του προσώπου είχε ικανοποιητικά αποτελέσματα, καθώς εντοπίστηκε η περιοχή του προσώπου και αποκόπηκε από την υπόλοιπη εικόνα επιτυχημένα. Μετέπειτα, έγινε μια προσπάθεια να οπτικοποιήσουμε τους χάρτες χαρακτηριστικών κάθε συνελικτικού επιπέδου 5.5.

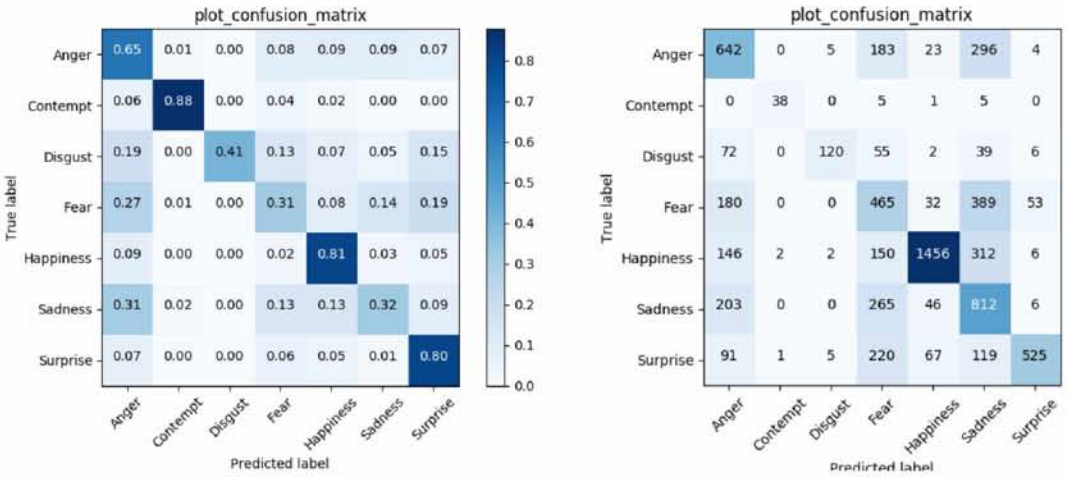
id	Επίπεδο	Διαστάσεις	Υπερπαράμετροι	Παράμετροι
1	Convolution2D_1	$48 \times 48 \times 32$	32, (3×3)	320
2	Activation	$48 \times 48 \times 32$	ReLU	
3	Convolution2D_2	$46 \times 46 \times 32$	32, (3×3)	9248
4	Activation	$46 \times 46 \times 32$	ReLU	
5	MaxPooling	$23 \times 23 \times 32$		
6	Dropout	$23 \times 23 \times 32$	0.5	
7	Convolution2D_3	$21 \times 21 \times 64$	64, (3×3)	18496
8	Activation	$21 \times 21 \times 64$	ReLU	
9	BatchNorm	$21 \times 21 \times 64$		256
10	MaxPooling	$10 \times 10 \times 64$		
11	Dropout	$10 \times 10 \times 64$	0.5	
12	Flatten	6400		
13	Dense	64		409664
14	Activation	64	ReLU	
15	Dropout	64	0.5	
16	Dense	7		455
17	Activation	7	Softmax	

ΠΙΝΑΚΑΣ 5.2: Αρχιτεκτονική του 1ου μοντέλου ΝΔΣ.

5.2.1 Αποτελέσματα 1ου μοντέλου

Το ΝΔΣ παρουσίασε μεγαλύτερη ακρίβεια στα αποτελέσματα σε σύγκριση με το baseline με ακρίβεια στο σύνολο επικύρωσης στο ποσοστό 63.14% και ακρίβεια στο σύνολο δοκιμής στο 60.31%. Τα αποτελέσματα προήλθαν από την συνάρτηση evaluate από το Keras και για τα αποτελέσματα ακρίβειας, ανάκλησης, F1 σκορ χρησιμοποιήθηκαν συναρτήσεις της sklearn.

Επιπλέον, τα αποτελέσματα οπτικοποιήθηκαν με την χρήση μητρώων σύγχυσης για την καλύτερη αντίληψη και κατανόηση τους. Υπάρχουν 2 πίνακες, εκ των οποίων ο ένας περιλαμβάνει τις πιθανότητες πρόβλεψης κάθε κλάσης και ο άλλος τις πραγματικές τιμές. Όπως φαίνεται και στους πίνακες 5.4 μπορούμε να παρατηρήσουμε πως το συναίσθημα της περιφρόνησης παρουσιάζει την μεγαλύτερη ακρίβεια ανάμεσα σε όλα τα συναισθήματα αλλά το δείγμα είναι πολύ μικρό για την συγκεκριμένη κλάση. Παρατηρείται επιπλέον πως τα συναισθήματα του φόβου και της λύπης παρουσιάζουν τις μεγαλύτερες απώλειες και παρατηρείται πως το πρόγραμμα συγχέει αυτά τα 2 συναισθήματα, καθώς και αυτό του θυμού με το συναίσθημα της λύπης.



Σχημα 5.4: Μητρώο σύγχυσης από το πρώτο μοντέλο ΝΔΣ με πραγματικές τιμές και μητρώο σύγχυσης με την πιθανότητα πρόβλεψης κάθε κλάσης.

	Ορθότητα	Απώλειες	Ακρίβεια	Ανάκληση	F1 σκορ
Σύνολο επικύρωσης	0.6314	0.9421			
Σύνολο δοκιμής	0.6031	0.9801	0.6030	0.6064	0.5772

Πινακας 5.3: Οι απώλειες, η ακρίβεια, η ανάκληση, F1 σκορ στο πρώτο ΝΔΣ, στο σύνολο επικύρωσης και στο σύνολο δοκιμής.

Μία ακόμα παράμετρος που εξετάστηκε είναι ο αλγόριθμος βελτιστοποίησης. Δοκιμάστηκαν οι αλγόριθμοι RMSprop, adadelta και adagrad στην υπάρχουσα αρχιτεκτονική. Παρατηρήθηκε πως ο αλγόριθμος adadelta σημείωσε τα μεγαλύτερα νούμερα ακρίβειας, ενώ ο αλγόριθμος adagrad άργησε να σημειώνει βήματα βελτίωσης στα πρώτα στάδια του.

rmsPROP	Ορθότητα	Απώλειες	Παράμετροι
Σύνολο επικύρωσης	0.6314	0.9558	lr=0.001, rho=0.9 ,
Σύνολο δοκιμής	0.6025	1.0516	epsilon=None , decay=0.0
Adadelta			
Σύνολο επικύρωσης	0.6342	0.9421	lr=1.0, rho=0.95
Σύνολο δοκιμής	0.6231	0.9801	epsilon=None, decay=0.0
Adagrad			
Σύνολο επικύρωσης	0.5931	1.0473	lr=0.01, epsilon=None,
Σύνολο δοκιμής	0.5865	1.0883	decay=0.0

Πινακας 5.4: Αποτελέσματα με βάση τον αλγόριθμο βελτιστοποίησης.

5.3 Τοπολογία 2ου μοντέλου ΝΔΣ και αποτελέσματα.

Το δεύτερο μοντέλο που υλοποιήθηκε βασίστηκε επίσης στις ΒΔ (CK+, FER2013). Για να ξεπεραστούν προβλήματα με το φαινόμενο υπερπροσαρμογής εφαρμόστηκε προσαύξηση των δεδομένων για την ενίσχυση της σταθερότητας του αλγορίθμου. Μετέπειτα, δημιουργήθηκε ένα καινούριο μοντέλο ΝΔΣ πιο σύνθετο, προσαρμοσμένο στα νέα δεδομένα.

Το νέο αυτό μοντέλο περιλαμβάνει 5 συνελικτικά επίπεδα. Το πρώτο επίπεδο χρησιμοποιεί ως συνάρτηση ενεργοποίησης την ReLU και αποτελείται από 32 φίλτρα με μέγεθος (5×5). Το δεύτερο συνελικτικό επίπεδο χρησιμοποιεί της ίδια συνάρτηση ενεργοποίησης και διαθέτει 32 φίλτρα (3×3) μεγέθους και ακολουθείται από ένα επίπεδο batch normalization, από ένα επίπεδο συγκέντρωσης διαστάσεων (2×2) καθώς και από ένα επίπεδο dropout με τιμή 0.5. Το τρίτο συνελικτικό επίπεδο αποτελείται από 64 φίλτρα με τον ίδιο σχηματισμό με πριν αλλά χωρίς επίπεδο συγκέντρωσης. Το τέταρτο συνελικτικό επίπεδο διαθέτει 64 φίλτρα και τον ίδιο σχηματισμό αλλά χωρίς επίπεδο dropout. Το πέμπτο επίπεδο περιλαμβάνει τον ίδιο σχηματισμό με το δεύτερο αλλά διαθέτει 128 συνελικτικά φίλτρα. Στο τέλος ακολουθούν πλήρως συνδεδεμένα επίπεδα που προκύπτουν από δύο επίπεδα Dense με αριθμό 1024 κόμβων και κάθε επίπεδο συνοδεύεται από ένα επίπεδο dropout με τιμή 0.5.

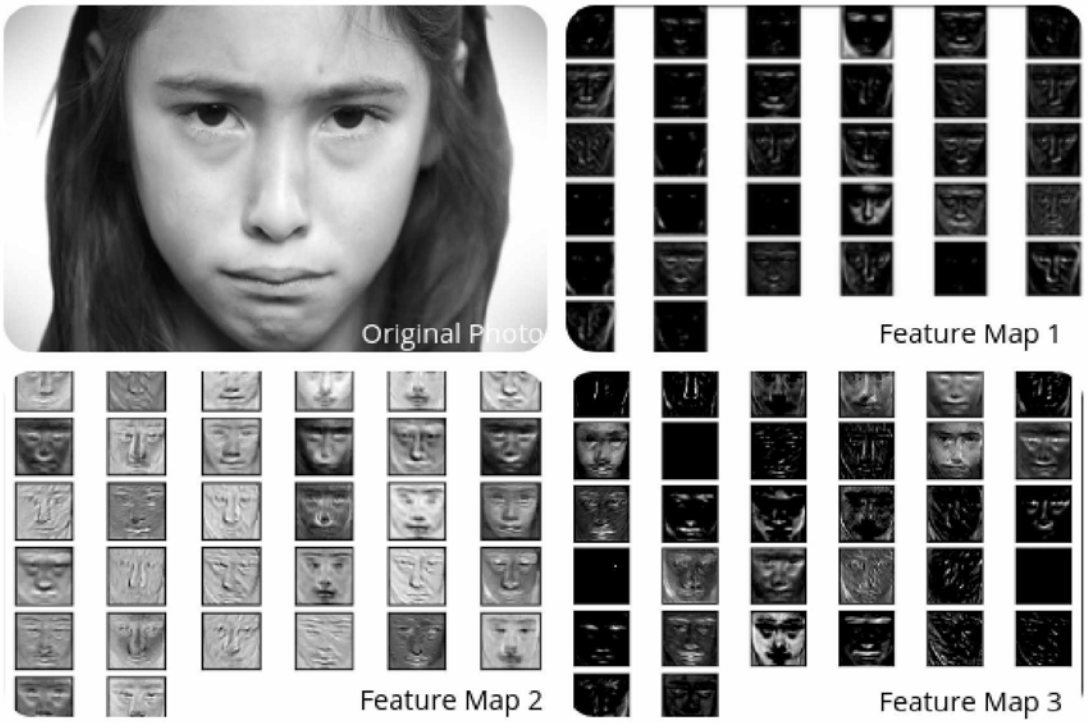
Παρακάτω παρουσιάζονται οι χάρτες χαρακτηριστικών για κάθε συνελικτικό επίπεδο, καθώς και οι τοπολογία του μοντέλου μέσα από έναν πίνακα. Η συγκεκριμένη αρχιτεκτονική περιλαμβάνει συνολικά 2.377.959 παραμέτρους από τις οποίες 2.377.383 είναι εκπαιδεύσιμες και 576 σταθερές.

Για την εκπαίδευση του μοντέλου χρησιμοποιήσαμε ως αλγόριθμο βελτιστοποίησης τον adadelta καθώς παρουσίασε τις καλύτερες επιδόσεις. Επιπλέον το μοντέλο εκπαιδεύτηκε σε batches = 256, epochs = 100.

id	Επίπεδο	Διαστάσεις	Υπερπαραμετροι	Παράμετροι
1	Convolution2D_1	$48 \times 48 \times 32$	$32, (5 \times 5)$	832
2	Activation	$48 \times 48 \times 32$	ReLU	
3	Convolution2D_2	$46 \times 46 \times 32$	$32, (3 \times 3)$	9248
4	Activation	$46 \times 46 \times 32$	ReLU	
5	BatchNorm	$46 \times 46 \times 32$		128
6	MaxPooling	$23 \times 23 \times 32$	2×2	
7	Dropout	$23 \times 23 \times 32$	0.5	
8	Convolution2D_3	$21 \times 21 \times 64$	$64, (3 \times 3)$	18496
9	Activation	$21 \times 21 \times 64$	ReLU	
10	BatchNorm	$21 \times 21 \times 64$		256
11	Dropout	$21 \times 21 \times 64$	0.5	
12	Convolution2D_4	$19 \times 19 \times 64$	$64, (3 \times 3)$	36928
13	Activation	$19 \times 19 \times 64$	ReLU	
14	BatchNorm	$19 \times 19 \times 64$		256
15	MaxPooling	$9 \times 9 \times 64$	2×2	
16	Convolution2D_5	$7 \times 7\times$ 128	$128, (3 \times 3)$	73856
17	Activation	$7 \times 7 \times 128$	ReLU	
18	BatchNorm	$7 \times 7 \times 128$		512
19	MaxPooling	$3 \times 3 \times 128$		
20	Dropout	$3 \times 3 \times 128$	0.5	
21	Flatten	1152		
22	Dense_1	1024	ReLU	1180672
23	Activation	1024		
24	Dense_2	1024		1049600
25	Activation	1024	ReLU	
26	Dropout	1024	0.5	
27	Dense_3	7		7175
28	Activation	7	Softmax	

Πινακας 5.5: Αρχιτεκτονική του 2ου μοντέλου ΝΔΣ.

Η προσπάθεια που έγινε στη συγκεκριμένη τοπολογία είχε ως στόχο την δημιουργία ενός ΝΔ με μεγαλύτερο βάθος, αλλά ταυτόχρονα να διατηρηθεί ο αριθμός των παραμέτρων σε μια μεσαία κλίμακα που θα επέτρεπε στην κάρτα γραφικών που χρησιμοποιείται να την υποστηρίξει. Επιπλέον, το μέγεθος των εικόνων περιορίστηκε μειώνοντας αισθητά τον όγκο της διαθέσιμης πληροφορίας για το ΝΔ.



Σχημα 5.5: Χάρτες χαρακτηριστικών μετά τα 3 πρώτα συνελικτικά επίπεδα.

5.3.1 Αποτελέσματα 2ου μοντέλου

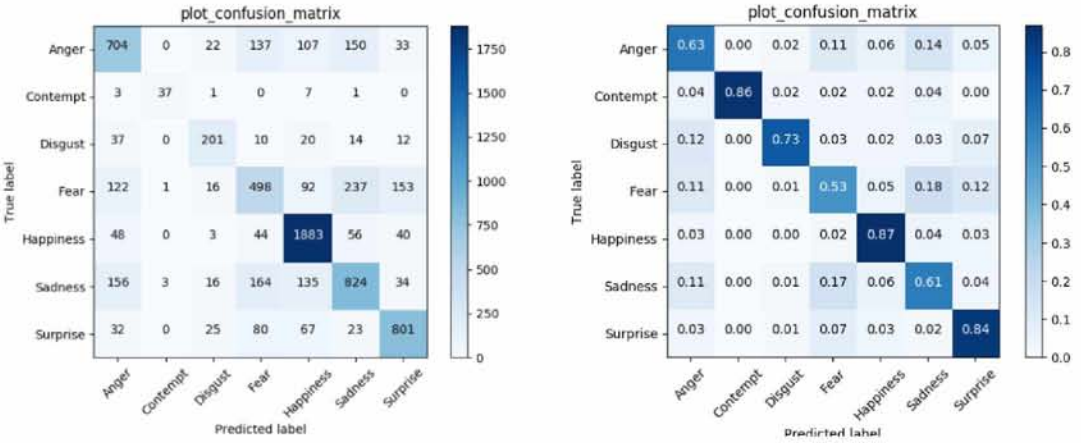
Στον πίνακα παρακάτω παρουσιάζονται τα ποσοστά ακριβείας και λάθους από το 2ο ΝΔΣ μοντέλο που κατασκευάστηκε. Η ακρίβεια του νέου δικτύου ανέρχεται στο 0.7099 στο σύνολο δοκιμής με την τιμή του λάθους στο 0.7908. Παρακάτω, μέσα από το μητρώο σύγχυσης εντοπίζονται τα συναισθήματα που μπερδεύουν περισσότερο το πρόγραμμα μας καθώς και αυτό που λαμβάνει τις καλύτερες προβλέψεις.

Το μητρώο σύγχυσης από το δεύτερο μοντέλο ΝΔΣ μαρτυρά τους συσχετισμούς μεταξύ των συναισθημάτων που μπερδεύουν το πρόγραμμα μας. Παρατηρώντας το

μητρώο σύγχυσης συμπεραίνουμε πως το συναίσθημα της χαράς αναγνωρίζεται με τα μεγαλύτερα ποσοστά ακρίβειας ανάμεσα σε όλα τα συναισθήματα. Τα συναισθήματα που μπερδεύουν περισσότερο το ΝΔ είναι ο θυμός, η λύπη και ο φόβος. Το συναίσθημα του φόβου αναγνωρίζεται λανθασμένα σε πολλές περιπτώσεις σε αυτό της λύπης καθώς και το αντίστροφο.

	Ορθότητα	Απώλειες	Ακρίβεια	Ανάκληση	F1 σκορ
Σύνολο επικύρωσης	0.7211	0.7323			
Σύνολο δοκιμής	0.7171	0.7704	0.7235	0.7234	0.7227

Πινακας 5.6: Οι απώλειες, η ακρίβεια, η ανάκληση, F1 σκορ στο δεύτερο ΝΔΣ, στο σύνολο επικύρωσης και στο σύνολο δοκιμής.



Σχημα 5.6: Μητρώο σύγχυσης από το δεύτερο μοντέλο ΝΔΣ με πραγματικές τιμές και μητρώο σύγχυσης με την πιθανότητα πρόβλεψης κάθε κλάσης.

5.4 Τοπολογία 3ου μοντέλου ΝΔΣ και αποτελέσματα

Στα προηγούμενα μοντέλα που δημιουργήθηκαν παρατηρήθηκε πως εξαιτίας του μικρού μεγέθους των εικόνων (48×48) και του βάθους του ΝΔ, ύστερα από ορισμένα επίπεδα συνέλιξης και συγκέντρωσης, το μέγεθος των δεδομένων συρρικνώνεται απότομα. Για αυτό τον λόγο προσθέσαμε πριν από το 2ο, 3ο και 4ο συνέλιξης από ένα

id	Επίπεδο	Διαστάσεις	Υπερπαράμετροι	Παράμετροι
1	Convolution2D_1	$48 \times 48 \times 32$	32, (5×5)	832
2	Activation	$48 \times 48 \times 32$	ReLU	
3	ZeroPadding	$50 \times 50 \times 32$	(1,1)	
4	Convolution2D_2	$48 \times 48 \times 32$	32, (3×3)	9248
5	Activation	$48 \times 48 \times 32$	ReLU	
6	BatchNorm	$48 \times 48 \times 32$		128
7	MaxPooling	$24 \times 24 \times 32$	2×2	
8	Dropout	$24 \times 24 \times 32$	0.5	
9	ZeroPadding	$26 \times 26 \times 32$	(1,1)	
10	Convolution2D_3	$24 \times 24 \times 64$	64, (3×3)	18496
11	Activation	$24 \times 24 \times 64$	ReLU	
12	BatchNorm	$24 \times 24 \times 64$		256
13	Dropout	$24 \times 24 \times 64$	0.5	
14	ZeroPadding	$26 \times 26 \times 64$	(1,1)	
15	Convolution2D_4	$24 \times 24 \times 64$	64, (3×3)	36928
16	Activation	$24 \times 24 \times 64$	ReLU	
17	BatchNorm	$24 \times 24 \times 64$		256
18	MaxPooling	$12 \times 12 \times 64$	2×2	
19	Convolution2D_5	$10 \times 10 \times 128$	128, (3×3)	73856
20	Activation	$10 \times 10 \times 128$	ReLU	
21	BatchNorm	$10 \times 10 \times 128$		512
22	MaxPooling	$5 \times 5 \times 128$		
23	Dropout	$5 \times 5 \times 128$	0.5	
24	Flatten	3200		
25	Dense_1	1024		3277824
26	Activation	1024	ReLU	
27	Dense_2	1024		627300
28	Activation	1024	ReLU	
29	Dropout	1024	0.5	
30	Dense_3	7		4291
31	Activation	7	Softmax	

ΠΙΝΑΚΑΣ 5.7: Αρχιτεκτονική του 3ου μοντέλου ΝΔΣ.

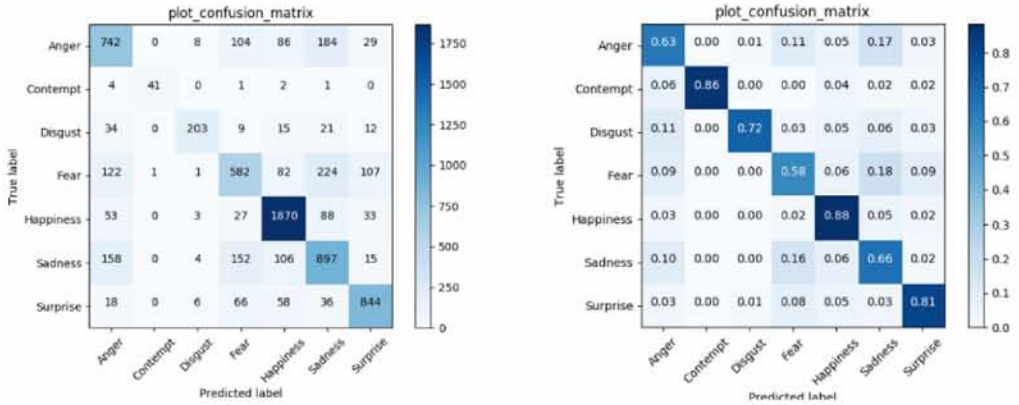
επίπεδο ZeroPadding, με μέγεθος (1×1). Με αυτό τον τρόπο διατηρείται το μέγεθος των δεδομένων ανέπαφο ύστερα από την πράξη της συνέλιξης. Ο αριθμός των παραμέτρων αυξήθηκε όπως είναι λογικό στις 4,049,351. Όλα τα υπόλοιπα στοιχεία παρέμειναν ίδια καθώς και οι παράμετροι εκπαίδευσης.

5.4.1 Αποτελέσματα 3ου μοντέλου

Από τα αποτελέσματα παρατηρούμε πως η ακρίβεια εμφανίζεται ελαφρώς ανεβασμένη στο 73.5%. Το ίδιο συμβαίνει και με τις απώλειες. Το συναίσθημα στο οποίο παρουσιάστηκε μεγαλύτερη βελτίωση είναι αυτό της λύπης αν και συνεχίζει να υπάρχει σύγχυση του συγκεκριμένου συναισθήματος με το συναίσθημα του φόβου και του θυμού.

	Ορθότητα	Απώλειες	Ακρίβεια	Ανάκληση	F1 σκορ
Σύνολο επικύρωσης	0.7347	0.7670			
Σύνολο δοκιμής	0.7312	0.7701	0.7750	0.7267	0.7471

Πινακας 5.8: Οι απώλειες, η ακρίβεια, η ανάκληση, F1 σκορ στο τρίτο ΝΔΣ, στο σύνολο επικύρωσης και στο σύνολο δοκιμής.



Σχημα 5.7: Μητρώο σύγχυσης από το τρίτο μοντέλο ΝΔΣ με πραγματικές τιμές και μητρώο σύγχυσης με την πιθανότητα πρόβλεψης κάθε κλάσης.

Κεφάλαιο 6

Συμπεράσματα και Μελλοντική Εργασία

6.1 Συμπεράσματα

Στα πλαίσια της πτυχιακής εργασίας με στόχο την αναγνώριση συναισθημάτων μέσα από τις εκφράσεις του προσώπου υλοποιήθηκε ένα μοντέλο ΝΔΣ το οποίο είναι αποτέλεσμα μια σειράς πειραμάτων, διορθώσεων και δοκιμών. Το αρχικό πόρισμα που εξάγεται από αυτή την διαδικασία είναι πως το συγκεκριμένο πρόβλημα μπορεί να επιλυθεί με την χρήση ΝΔΣ σε ικανοποιητικό βαθμό και να ξεπεράσει τις επιδόσεις άλλων αλγορίθμων όπως των ΜΔΣ που χρησιμοποιήθηκαν ως baseline, παρά τους περιορισμούς που υπάρχουν σε ζητήματα μνήμης. Επιπλέον, παρατηρείται πως αρκετοί παράμετροι μπορούν να επηρεάσουν την ακρίβεια των αποτελεσμάτων και να διαφοροποιήσουν το πρόβλημα και για αυτό τον λόγο θα πρέπει να πραγματοποιηθεί προεπεξεργασία πριν την είσοδο των δεδομένων στο ΝΔΣ. Το μέγεθος των εικόνων, ο φωτισμός και οι συνθήκες κατά τις οποίες δημιουργούνται τα δεδομένα ελέγχου αποτελούν βασικές παραμέτρους. Κατά την διάρκεια υλοποίησης της εργασίας, δοκιμάστηκαν διάφοροι αλγόριθμοι βελτιστοποίησης και ο αλγόριθμος adadelta παρουσίασε την βέλτιστη επίδοση. Τέλος, η προσθήκη των επιπέδων batch normalization και zero padding συνέβαλε ουσιαστικά στην βελτίωση των αποτελεσμάτων καθώς επίσης και η χρήση της τεχνικής της εξισορρόπησης ιστογράμματος. Η προσέγγιση που ακολουθήθηκε στην τρίτη υλοποίηση, η οποία απέδωσε και την μεγαλύτερη ακρίβεια, θα μπορούσε να αποφέρει ακόμα πιο μεγάλη ακρίβεια με την χρήση και άλλων αλγορίθμων ΝΔ, καθώς μέσα από τα ΝΔΣ δεν εκμεταλλεύεται η

επιπλέον πληροφορία που μπορεί να παρέχει η συνεχόμενη ροή εικόνων μέσα από ένα βίντεο.

6.2 Μελλοντική εργασία

Σε αυτή την διπλωματική γίνεται μια προσπάθεια να επιλυθεί το ζήτημα της αναγνώρισης συναισθημάτων μέσα από βίντεο με την χρήση ΝΔΣ. Σίγουρα, πολλές βελτιστοποιήσεις, ρυθμίσεις και ιδέες για αυτή την εργασία αποτελούν δουλειά για το μέλλον εξαιτίας της έλλειψης χρόνου. Η μελλοντική δουλειά περιλαμβάνει λοιπόν βαθύτερη ανάλυση από συγκεκριμένους μηχανισμούς και μοντέλα ΝΔ όπως τα ΝΔΑ ή συνδυαστικά μοντέλα με χρήση ΝΔΣ με ΝΔΑ και LSTM (long short term memory) ή άλλες τεχνικές προεπεξεργασίας της εικόνας ικανές να επιφέρουν ακόμα μεγαλύτερη ακρίβεια στο δίκτυο. Επιπλέον, ένας μελλοντικός στόχος για την εργασία αυτή, είναι ο εμπλουτισμός της με δεδομένα που περιέχουν εικόνες που έχουν τραβηχτεί σε ελεύθερο περιβάλλον και όχι αποκλειστικά σε ειδικά διαμορφωμένα εργαστήρια, καθώς επίσης και εικόνες που περιλαμβάνουν περισσότερα από ένα πρόσωπα. Αυτή η ενέργεια θα προσδώσει στο δίκτυο μεγαλύτερη ευελιξία και δυνατότητα να αναγνωρίζει πιο πολύπλοκα μοτίβα. Επιπρόσθετα, υπάρχει η σκέψη να κατασκευαστεί ένας καινούργιος ταξινομητής αναγνώρισης προσώπου που θα έχει μεγαλύτερη ακρίβεια. Τέλος, αυτή η υλοποίηση θα ήταν ενδιαφέρον να προσαρμοστεί σε μία φιλική προς τον χρήστη εφαρμογή που θα χρησιμοποιηθεί για διάφορους σκοπούς.

Βιβλιογραφία

[1] Paul Ekman, Wallace V. Friesen, Maureen O'Sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E. Ricci-Bitti, Klaus Scherer, Masatoshi Tomita, and Athanase Tzavaras. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4):712–717, 1987. ISSN 1939-1315. doi: 10.1037/0022-3514.53.4.712.

[2] Dan C. Cireşan, Ueli Meier, Jonathan Masci, Luca M. Gambardella, and Jürgen Schmidhuber. Flexible, high performance convolutional neural networks for image classification. *IJCAI International Joint Conference on Artificial Intelligence*, pages 1237–1242, 2011. ISSN 10450823. doi: 10.5591/978-1-57735-516-8/IJCAI11-210.

[3] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: A survey. *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*, 41(6):765–781, 2011. ISSN 1094-6977. doi: 10.1109/TSMCC.2011.2118750.

[4] Dewi Yanti Liliana, Rahmat Widyanto, and T Basaruddin. Human emotion recognition based on active appearance model and semi-supervised fuzzy c-means. In *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 439–445, 2016.

[5] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 94–101, June 2010.

[6] Gil Levi and Tal Hassner. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 503–510, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3912-4. doi: 10.1145/2818346.2830587.

[7] Veena Mayya, Radhika M. Pai, and M.M. Manohara Pai. Automatic facial expression recognition using dcnn. *Procedia Computer Science*, 93:453 – 461, 2016. ISSN 1877-0509. Proceedings of the 6th International Conference on Advances in Computing and Communications.

[8] Gary McKeown, Michel F. Valstar, Roderick Cowie, and Maja Pantic. The SEMAINE corpus of emotionally coloured character interactions. In *2010 IEEE International Conference on Multimedia and Expo*, pages 1079–1084, July 2010. doi: 10.1109/ICME.2010.5583006.

[9] Olivier Martin, Irene Kotsia, Benoit M. Macq, and Ioannis Pitas. The enterface'05 audio-visual emotion database. In Roger S. Barga and Xiaofang Zhou, editors, *ICDE Workshops*, page 8. IEEE Computer Society, 2006. ISBN 0-7695-2571-7.

[10] Joseph Weizenbaum. ELIZA, A computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1): 36–45, January 1966. ISSN 0001-0782. doi: 10.1145/365153.365168.

[11] Nancy Alvarado. Arousal and valence in the direct scaling of emotional response to film clips. *Motivation and Emotion*, 21(4):323–348, Dec 1997. ISSN 1573-6644. doi: 10.1023/A:1024484306654.

[12] Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 64:59 – 63, 2015. ISSN 0893-6080.

[13] Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat. Web-based database for facial expression analysis. In *2005 IEEE International Conference on Multimedia and Expo*, pages 5–22, July 2005. doi: 10.1109/ICME.2005.1521424.

[14] Geoffrey E Hinton Josh M Susskind, Adam K Anderson. Toronto face database. *Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech.Rep*, 2010.

[15] Takeo Kanade and Jeffrey F. Cohn. Comprehensive database for facial expression analysis. *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000. ISSN 0-7695-0580-5. doi: 10.1109/AFGR.2000.840611.

[16] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 2106–2112, Nov 2011. doi: 10.1109/ICCVW.2011.6130508.

[17] Ian Sneddon, Margaret McRorie, Gary McKeown, and Jennifer Hanratty. The belfast induced natural emotion database. *IEEE Transactions on Affective Computing*, 3(1):32–41, Jan 2012. ISSN 1949-3045. doi: 10.1109/T-AFFC.2011.26.

[18] Tanja Bänziger, Marcello Mortillaro, and Klaus Scherer. Introducing the geneva multimodal expression corpus for experimental research on emotion perception. 12:1161–79, Nov 2011.

[19] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image Vision Computing*, 28(5):807–813, May 2010. ISSN 0262-8856. doi: 10.1016/j.imavis.2009.08.002.

[20] Marian Stewart Bartlett, Gwen Littlewort, Mark G. Frank, Claudia Lainscsek, Ian R. Fasel, and Javier R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1:22–35, 2006.

[21] Carlos Busso and Shrikanth S. Narayanan. Recording audio-visual emotional databases from actors: a closer look. In *Second International Workshop on*

Emotion: Corpora for Research on Emotion and Affect, International conference on Language Resources and Evaluation (LREC 2008), pages 17–22, Marrakech, Morocco, May 2008.

[22] Ronak Kosti, Jose M. Alvarez, Adria Recasens, and Agata Lapedriza. Emotic: Emotions in context dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2309–2317, July 2017. doi: 10.1109/CVPRW.2017.285.

[23] Michael Grimm, Kristian Kroschel, and Shrikanth Narayanan. The vera am mittag german audio-visual emotional speech database. In *2008 IEEE International Conference on Multimedia and Expo*, pages 865–868, June 2008. doi: 10.1109/ICME.2008.4607572.

[24] Satish Anila and Nanjundappan Devarajan. Preprocessing technique for face recognition. *Global Journal of Computer Science and Technology Graphics & Vision*, 12(11):13–18, 2012.

[25] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *In Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR. URL http://proceedings.mlr.press/v37/ioffe15.html.

[26] Principal component analysis. https://en.wikipedia.org/wiki. Accessed: 2018-05-16.

[27] Alexander Ilin and Tapani Raiko. Practical approaches to principal component analysis in the presence of missing values. *Journal of Machine Learning Research*, 11:1957–2000, 2010. ISSN 15324435. doi: 10.1109/TPAMI.2010.46.

[28] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *CoRR*, abs/1712.04621, 2017.

[29] François Chollet et al. Keras. https://github.com/keras-team/keras, 2015.

[30] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. *In Proceedings of the 2001 IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1: I–511–I–518, 2001. ISSN 1063-6919. doi: 10.1109/CVPR.2001.990517.

[31] M. Martínez-Zarzuela, F. J. Díaz-Pernas, M. Antón-Rodríguez, F. Perozo-Rondón, and D. González-Ortega. Adaboost face detection on the gpu using haar-like features. In José Manuel Ferrández, José Ramón Álvarez Sánchez, Félix de la Paz, and F. Javier Toledo, editors, *New Challenges on Bioinspired Applications*, pages 333–342, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-21326-7.

[32] David H. Hubel and Torsten N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195: 215–243, 1968.

[33] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.

[34] Anirud Pande and Rohit Chandna. Matrix convolution using parallel programming. *International Journal of Science and Research (IJSR)*, 2(7):286–291, 2013.

[35] Course stanford 231, convolutional neural networks for visual recognition. http://cs231n.stanford.edu. Accessed: 2018-05-16.

[36] Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In Konstantinos Diamantaras, Wlodek Duch, and Lazaros S. Iliadis, editors, *Artificial Neural Networks – ICANN 2010*, pages 92–101, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-15825-4.

[37] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2):301–320, 2005. ISSN 13697412. doi: 10.1111/j.1467-9868.2005.00503.x.

[38] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. ISSN 15337928. doi: 10.1214/12-AOS1000.

[39] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.

[40] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[41] Geoffrey E. Hinton, Nitish Srivastava, and Kevin Swersky. Lecture 6a-overview of mini-batch gradient descent. *COURSERA: Neural Networks for Machine Learning*, page 31, 2012.

[42] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012. ISSN 10495258. doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

[43] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions. *arXiv:1409.4842*, 2014. ISSN 10636919. doi: 10.1109/CVPR.2015.7298594.

[44] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2015. URL http://arxiv.org/abs/1409.1556.